SEVERITY GRADING AND EARLY DETECTION OF ALZHEIMER'S DISEASE
THROUGH TRANSFER LEARNING

by

SAEED MOHAMMED S ALQAHTANI

A dissertation submitted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY IN ELECTRICAL AND COMPUTER ENGINEERING

2025

Oakland University
Rochester, Michigan

Doctoral Advisory Committee:

Mohamed Zohdy, Ph.D., Chair
Subramaniam Ganesan, Ph.D.
Christopher Kobus, Ph.D.
Suzan ElSayed, Ph.D.

To my mother, father and siblings

ACKNOWLEDGMENTS

I would like to express my wholehearted gratitude to all of those who helped, encouraged, and supported me throughout my Ph.D. journey. First and foremost, I would like to express my utter appreciation and gratefulness to my advisor and committee chair Dr. Mohamed Zohdy for his relentless guidance, support, and encouragement as they were instrumental in my research development. I also would like to express my earnest appreciation to my dissertation committee members, Dr. Subramaniam Ganesan, Dr. Chris Kobus, and Dr. Suzan Elsayed; their unwavering support and encouragement, and constructive criticism have elevated the caliber of this work immeasurably.

I also would like to thank Dr. Louay Chamra, Dean of the College of Engineering and Computer Science, Dr. Qian (Beth) Zou, Associate Dean of the College of Engineering and Computer Science, and special thanks to Dr. Osamah A. Rawashdeh, Chair of the Electrical and Computer Engineering Department for their support through my journey at Oakland University.

I owe immense gratitude to my family and friends. To my parents, thank you for your boundless support and love and for cultivating critical thinking in us. This moment marks not the end, but the beginning of an intellectual journey with many significant milestones yet to be achieved.

Saeed Mohammed S Alqahtani

ABSTRACT

SEVERITY GRADING AND EARLY DETECTION OF ALZHEIMER'S DISEASE THROUGH TRANSFER LEARNING

by

SAEED MOHAMMED S ALQAHTANI

Adviser: Mohamed Zohdy, Ph.D.

Alzheimer's disease (AD) is a neurological disorder that predominantly affects individuals aged 65 and older. It is one of the primary causes of dementia, and it contributes significantly and progressively to impairing and destroying brain cells. Recently, efforts to mitigate the impact of AD have focused with particular emphasis on early detection through computer aided diagnosis (CAD) tools. This study aims to develop deep learning models for the early detection and classification of AD cases into four categories: non-demented, moderate-demented, mild-demented, and very mild-demented. Using Transfer Learning technique (TL), several models were implemented including AlexNet, ResNet-50, GoogleNet (InceptionV3), and SqueezeNet, by leveraging magnetic resonance images (MRI) and applying image augmentation techniques. A total of 12,800 images across the four classifications that were preprocessed to ensure balance and meet the specific requirements of each model. The dataset was split into 80% for training and 20% for testing. AlexNet achieved an average accuracy of 98.05%, GoogleNet (InceptionV3) reached 97.80%, ResNet-50 attained 91.11%, and SqueezeNet 86.37%. The use of transfer learning method addresses data limitations, allowing effective model training without the need for building from scratch, thereby enhancing the potential for early and accurate diagnosis of Alzheimer's disease [1].

TABLE OF CONTENTS

TABLE OF CONTENTS—Continued

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AD	Alzheimer’s Disease
SPECT	Single-Photon Emission Computed Tomography
PET	Positron Emission Tomography
MRI	Magnetic Resonance Images
fMRI	functional Magnetic Resonance Images
CT	Computed Tomography
MCI	Mild Cognitive Impairment
DL	Deep Learning
NN	Neural Network
SVM	Support Vector Machines
CNN	Convolution Neural Network
AI	Artificial Intelligence
ML	Machine Learning Algorithm
ReLU	Rectified Linea Unit
GELU	Gaussian Error Linear Unit
TL	Transfer Learning
SGD momentum	Stochastic Gradient Descent with Momentum
RGB	Red, Green, and Blue colors
SMOTE	Synthetic Minority Oversampling Technique
GANs	Generative Adversarial Networks
NST	Neural Style Transfer

ADNI	Alzheimer's Disease Neuroimaging Initiative
SGDM	Stochastic Gradient Descent with Momentum
SOMs	Self-Organizing Maps
AutoML	Automated Machine Learning

CHAPTER ONE

INTRODUCTION AND BACKGROUND

1.1 Alzheimer's Disease

Alzheimer's disease (AD) is a neurological disorder that primarily affects and targets the brain and nervous system. It is a chronic and irreversible condition that disrupts essential cognitive functions like a person's abilities to carry out basic memory, thinking, and social tasks. As one of the most prevalent forms of dementia, AD initially manifests through symptoms such as memory loss, language skills challenges, and struggles with thinking and self-management [2]. Over time, the disease progressively erodes memory and encumbers the ability to perform routine daily activities. As AD progresses and cognitive functions continue to decline, patients face increasing difficulties in maintaining an independent and normal life. Such development places an incredible strain on patients' independence leading them to rely heavily on family members and caregivers for their basic needs and survival as the disease advances [1].

Alzheimer's disease is one of the most common and recognized forms of dementia, particularly among individuals aged 65 and older. Projections indicate that by 2025, 1 in every 85 people will be diagnosed with Alzheimer's, and that number is expected to double over the subsequent 20 years. According to the Alzheimer's Association report, age is one of the most significant risk factors for Alzheimer's disease. The probability of being affected by AD rises with age 65 to 74 years, 75 to 84 years by 5.0% to 13.1%, and 85 years and older by 33.2%. Also, the Alzheimer's Association estimates that in the year 2022, approximately 6.5 million Americans aged 65 and older were living with Alzheimer’s disease as illustrated in the figure below, figure 1.1 [2].

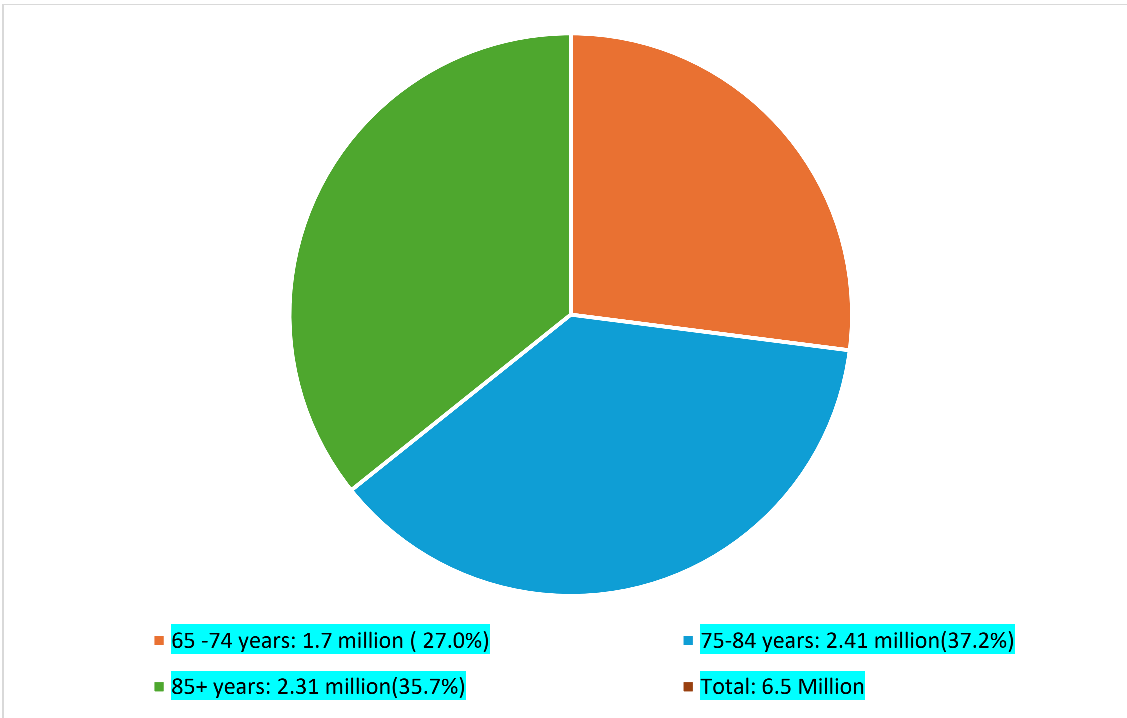

Figure 1.1 Cases of People 65+ years with Alzheimer’s in the U.S 2022 [2].

According to the Alzheimer’s Association report. 2022 Alzheimer’s disease facts and figures, it’s projected that every state in the United States of America to observe an increase at least a 6.7% rise in the number of individuals with Alzheimer’s between 2020 and 2025. Table1.1 provides the prevalence estimates for 2020 and 202, in conjunction with the difference transpiring during this period. These predicted increases in Alzheimer's cases are primarily driven by anticipated shift in the population aged 65 and older across states, particularly the number of individuals at specific ages (e.g., 66, 67, etc.). The Southeast and West regions are predicted to experience the highest percentage growth in Alzheimer’s cases between 2020 and 2025, mainly due to shifts in the age demographic of their population over this period. These elevating numbers will substantially affect each state’s healthcare system and the Medicaid program which plays a critical role in financing long term care and support services for a significant number of adults with dementia including about a quarter of Medicare beneficiaries with Alzheimer's or other dementias. It is imperative to emphasize that regional patterns of current and future encumber do not account for potential future variations for other dementia risk factors such as midlife hypertension and diabetes, across different regions and states.

Table 1.1 Alzheimer's Cases Projection in the United States of America in 2025 by state [2].

State	**2020 (in thousands)**	**2025 (in thousands)**	**Percentage Increase 2020-2025**
Alabama	96	110	14.6%
Alaska	8.5	11	29.4%
Arizona	150	200	33.3%
Arkansas	58	67	15.5%
California	690	840	21.7%
Colorado	76	92	21.1%
Connecticut	80	91	13.8%
Delaware	19	23	21.1%

Table 1.1—Continued Alzheimer's Cases Projection in the United States of America in 2025 by state [2].

State	2020 (in thousands)	2025 (in thousands)	Percentage Increase 2020-2025
District of Columbia	8.9	9	1.1%
Florida	580	720	24.1%
Georgia	150	190	26.7%
Hawaii	29	35	20.7%
Idaho	27	33	22.2%
Illinois	230	260	13.0%
Indiana	110	130	18.2%
Iowa	66	73	10.6%
Kansas	55	62	12.7%
Kentucky	75	86	14.7%
Louisiana	92	110	19.6%
Maine	29	35	20.7%
Maryland	110	130	18.2%
Massachusetts	130	150	15.4%
Michigan	190	220	15.8%

Table 1.1—Continued Alzheimer's Cases Projection in the United States of America in 2025 by state [2].

State	2020 (in thousands)	2025 (in thousands)	Percentage Increase 2020-2025
Minnesota	99	120	21.2%
Mississippi	57	65	14.0%
Missouri	120	130	8.3%
Montana	22	27	22.7%
Nebraska	35	40	14.3%
Nevada	49	64	30.6%
New Hampshire	26	32	23.1%
New Jersey	190	210	10.5%
New Mexico	43	53	23.3%
New York	410	460	12.2%
North Carolina	180	210	16.7%
North Dakota	15	16	6.7%
Ohio	220	250	13.6%
Oklahoma	67	76	13.4%
Oregon	69	84	21.7%

Table 1.1—Continued Alzheimer's Cases Projection in the United States of America in 2025 by state [2].

State	2020 (in thousands)	2025 (in thousands)	Percentage Increase 2020-2025
Pennsylvania	280	320	14.3%
Rhode Island	24	27	12.5%
South Carolina	95	120	26.3%
South Dakota	18	20	11.1%
Tennessee	120	140	16.7%
Texas	400	490	22.5%
Utah	34	42	23.5%
Vermont	13	17	30.8%
Virginia	150	190	26.7%
Washington	120	140	16.7%
West Virginia	39	44	12.8%
Wisconsin	120	130	8.3%
Wyoming	10	13	30.0%

In the twenty years leading up to the COVID-19 pandemic, official reported data show a significant rise in deaths attributed to Alzheimer’s disease while mortality rates from other major causes either declined substantially or remained relatively stable. From

2000 to 2019, the number of deaths from Alzheimer's disease surged by 145.2% as documented on death certificates, more than doubled. On the contrary, deaths from heart disease which is the leading cause of mortality in the United States, decreased by 7.3% (see Figure 1.2 below).

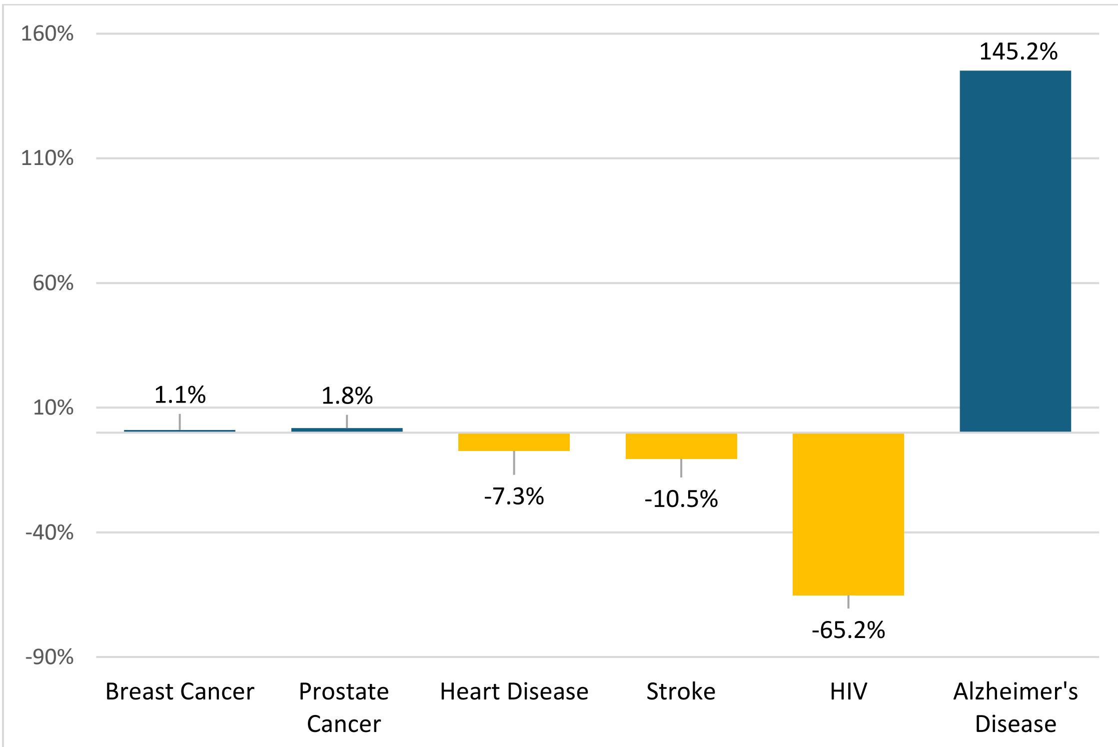

Figure 1.2 Rate of change in various mortality causes for all ages between 2000 and 2019, derived from National Center for Health Statistics data [2].

Alzheimer's disease was officially identified as the sixth leading cause of mortality in the United States in 2019, and the seventh leading cause of mortality in 2020 and 2021. However, for individuals aged 65 and older, Alzheimer's remains the fifth leading cause of death even though the actual true number of fatalities might be substantially higher than what is reflected in the official reports. Those living with Alzheimer's endure years of progressive deterioration prominently with poor health, severe disability, and debilitating symptoms before capitulating to AD. According to Alzheimer's Association 2022 report, 121,499 deaths in 2019 were ascribe to Alzheimer's disease; conversely, the report also highlights that these numbers likely underestimate deaths from Alzheimer's as the cause of death is often inaccurately reported. AD is frequently associated with other serious illness and health complications, especially among older adults such as pneumonia as severe dementia can lead to additional issues like swallowing difficulties, immobility, and malnutrition, all of which raise the probability of developing acute conditions and eventual death (comorbidity). The estimated total cost of long-term medical care for dementia patients in 2021 reached approximately USD 290 billion [14]. While there is no known cure currently, early detection is critical as it allows for interventions which may slow down the progression of the disease [2]. Detecting AD before clinical symptoms appear is vital for initiating early treatment and improving outcomes [1].

To precisely identify Alzheimer’s disease (AD) and distinguish its various stages from healthy controls, multiclass decision-making system is essential. Since AD often goes undetected until it is too late for effective intervention, especially without the involvement of specialists or advanced diagnostic skills, the ability to differentiate AD from mild cognitive impairment (MCI) or cognitively healthy people becomes critically important [15].

Despite the fact that the causes of Alzheimer’s disease remain not fully understood, the disease transpires when abnormalities in protein function cause damage to nerve cells, resulting in a breakdown of neural connectivity. Diagnosing AD require the presence of two specific irregularities: extracellular amyloid plaques and intracellular neurofibrillary tangles [15]. In the first abnormality, amyloid plaques accumulate on the exterior of neurons, interfering with their ability to communicate. In the second abnormality, neurofibrillary tangles, made up of filamentous tau proteins, further impair neural connectivity and are associated with the death of neurons [16]. These amyloid plaques and neurofibrillary tangles have long been recognized as key biomarkers and frequently utilized to assist in the diagnosis of AD [17].

In the field of imaging diagnostics, various types of imaging techniques are used to accurately diagnose Alzheimer's disease (AD) such as structural and functional imaging methods. SPECT, Single-photon emission computed tomography, is one such imaging technology used in AD diagnosis. Functional imaging examples include PET, positron emission tomography, and fMRI, functional magnetic resonance imaging, while structural imaging examples include MRI, magnetic resonance imaging and CT. A CT scan, which stands for computed tomography, utilizes multiple X-ray scans taken from different angles to create detailed images of the targeted organ. In contrast, MRI employs radio waves and electromagnetic fields to generate high-resolution images of the organ. Alternatively, PET and fMRI rely on molecular imaging techniques, which involve the use of radiotracers to detect chemical and cellular changes associated with specific diseases or conditions. Given the limited accessibility of functional imaging and the restricted capabilities of CT, MRI has become ubiquitously utilized due to its availability, efficiency, and versatility [2], [14]. Compared to other imaging practices, MRI stands out for its cost-effectiveness, superior diagnostic capabilities, and the absence of ionizing radiation which some people may be allergic to. It is applicable across all stages of Alzheimer's disease and can effectively differentiate between healthy individuals and the unwell. Furthermore, MRI is valuable for detecting mild cognitive impairment (MCI) and Alzheimer's disease (AD), as well as for monitoring the progression from MCI to AD. It also plays a vital role in secernating between different types of dementia, such as frontotemporal dementia and Alzheimer's, which may present similar clinical symptoms [18], [19].

1.2. Deep Learning

Neural networks' architectures were originally inspired by the human brain which subsequently inspired the architecture of deep learning. A neural network comprises of intertwined layered nodes or neurons and there are three main layers: input layer, hidden layer, output layer. As the previous order suggests, the hidden layer is the layer that comes between the input layer and the output layer which contains numerous numbers of neurons. Hundreds or even thousands of hidden layers could be In a deep learning model [3], [20]. In machine learning, deep learning is a subset which characterized by characterized by numerous layers and parameters, with most deep learning system employing neural network architectures. Ordinarily, deep learning demands a substantial volume of data. Once the dataset is collected and accurately labeled, it can advance to the deep learning phase [21], [22]. The most frequently employed deep learning models are training from scratch, transfer learning, and feature extraction. Training from scratch is the least favored method of the three as it necessitates constructing the model with meticulous attention to every detail along trial and error, though it is highly effective for novel applications. Feature extraction is the second preferred technique, requiring specific expertise to utilize the network as a feature extractor. Once the feature extraction process is complete, models like SVM, support vector machine, can be used for classification. ultimately, Transfer Learning is the most widely adopted method in deep learning application, involving preprocessing and fine-tuning of an existing pretrained network or model such as AlexNet and GoogLeNet [1], [23], [24].

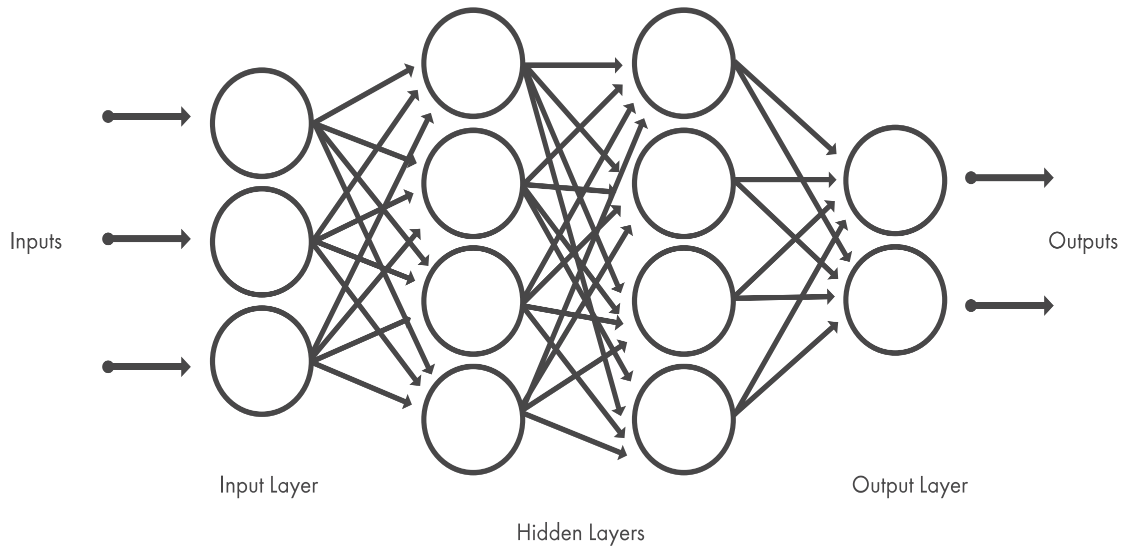

Figure 1.3 fundamental illustration of deep network architecture [3].

<u>1.2.1 Convolution Neural Network</u>

A driven computer aided diagnosis (CAD) systems by artificial intelligence have emerged as a vastly admired and viable tool in the field of medicine, mainly because of its cost effectiveness and transparent decision-making process [25]. An important area within AI is deep neural networks-based machine learning (ML) algorithm, where computers are enabled to analyze datasets by identifying similarities and differences to establish meaningful connections for decision-making, all without requiring direct human intervention. The majority of advancements in machine learning algorithms were initially aimed at surpassing human capabilities, specifically during the late 1990s, however, these systems have yet to achieve fully satisfactory functionality [1], [26].

A convolutional neural network (CNN) was initially developed through the inspiration from the visual processing mechanism of the human brain [27]. However, the innovation and evolution didn't halt at this stage, it advanced further to encompass deep convolutional neural networks. The deep convolution neural networks have achieved astonishing and powerful levels of learning capabilities particularly because they employ multiple feature extraction methods that intuitively recognize patterns and identify unique features. In recent years, computing power has grown considerably, driven by the development of more robust processors, which have significantly contributed to the progress and refinement of research and development in convolutional neural networks. One of the most distinctive characteristics of CNNs is their hierarchical architecture. The structure of a CNN is composed of several interconnected layers, including Convolution, Pooling, Activation, Normalization, and Fully Connected layers. The output of the convolution layer, often referred to as CONV, is generated by dividing the input image into smaller segments and applying weighted convolutions to these segments. Essentially, CONV layers are responsible for extracting key features and characteristics from the input images [26].

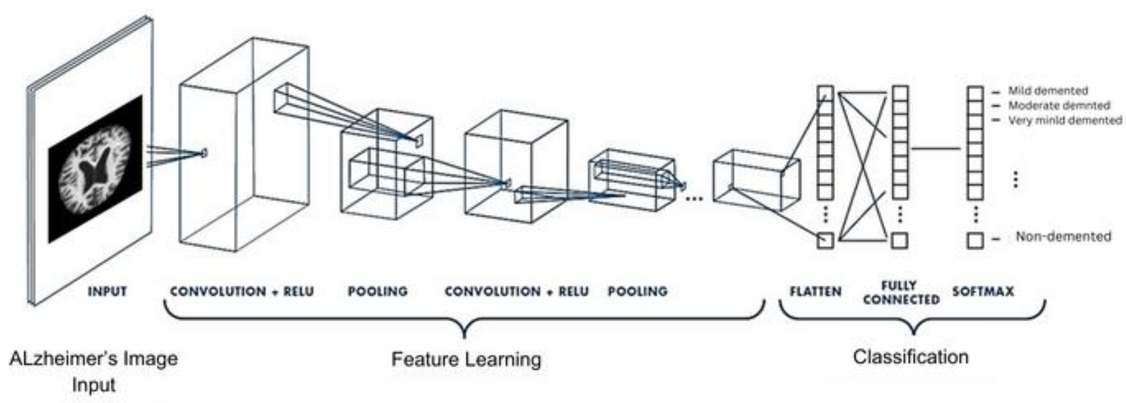

Figure 1.4 General illustration of Convolution Neural Networks architecture [4], [5].

Convolutional neural networks (CNNs) consist of layers designed to extract local information from an input image through convolutional operations such as edges. Each node within a convolutional layer is connected to a small number of neurons that have spatial relationships. The weights of these connections are shared across all nodes in the convolutional layer which in turns enable the detection of the same local feature throughout the entire input image. Each set of shared weights is referred to as a convolution kernel. In the interest of reducing computational complexity, a pooling layer follows every sequence of convolution layers. Among the various types of pooling layers, the max pooling layer is the most known and used. The max pooling layer reduces the dimensionality of feature maps by selecting the maximum feature response within neighboring regions, thereby preserving the most prominent features [28], [29]. The critical role that max pooling layer plays in CNN operation is reducing the size reducing the spatial dimensions of feature maps while preserving the most significant features. The process is done by gliding a fixed window, typically 2x2 or 3x3, across the input feature map and choosing the maximum value within each window. This donwsampling technique reduces the computational complexity as previously mentioned while simultaneously enhancing translation variance thereby reducing the network's sensitivity to minor shifts or distortions in the input data [30]. It’s known to be very effective in retaining the most activated features, hence, why it’s most favored in image classification and object detection over other approaches in pooling such as average pooling [31]. Correspondingly, the activation layer is an essential part of CNN as in introduces non-linearity into the model which in turn allows it to learn complex patters and relationships in the data. If the activation layer was not part of CNN, it would only performer linear transformations which it would drastically hinders its ability to model intricate structures within a dataset like images [32]. One of the most commonly used layers in activation layers in rectified Linea Unit (ReLU) because of its simplicity and effectiveness in mitigating the vanishing gradient issue. Nonetheless, ReLU is prone to what is known as “dying ReLU” where neurons become inactive and produce zero output for all input [33], [34]. To counteract this phenomenon, variants such as Parametric ReLU and Leaky ReLU were developed which allows a small gradient for negative inputs [32], [33]. recent developments in activation layer have led to introducing activation functions such as Swish and GELU (Gaussian Error Linear Unit) where they have exhibited outstanding performance in certain tasks by offering smoother gradient and improved generalization [35], [36]. The fully connected layer, also referred to as dense layer, is another part of conventional neural networks serving in the final stage in my architectures to carry out classification or regression tasks. In a fully connected layer, each neuron is linked to every neuron in the previous layer. This structure allows the network to learn complex, global relationships by integrating features extracted from prior layers [30]. Although fully connected layers are tremendously effective, they coincide with significant computational costs and susceptibility to overfitting, especially in deep networks characterized by an abundance of parameters [37]. Therefore, techniques such as dropout, weight regularization, and batch normalization were developed to mitigate such obstacles [38], [39]. Contemporary studies have aimed at enhancing fully connected layers by employing sparse connectivity and dynamic architecture which adaptively prune redundant connections during the training process [40], [41]. Additionally, incorporating fully connected layers with transformers and graph neural networks has expanded their applications to domains like natural language processing and graph-based learning [42], [43]. Notwithstanding the advent of the alternative to fully connected layers, it remains prevalent and essential in countless deep learning models, particularly in tasks necessitating high dimensional feature mapping [44], [45]. In the output layer, the softmax layer is often utilized for multiclass classification tasks since it converts raw scores to probabilities [24], [30]. The proposed approach employs deep learning models, including AlexNet, ResNet-50, and GoogleNet (InceptionV3), utilizing transfer learning (TL) for early detection and accurate classification of cases based on magnetic resonance images (MRI).

1.3. Transfer Learning

Transfer Learning is a technique used in deep learning where a previously trained network has its knowledge transferred over to perform a new task that may or may not be related to the previously attained knowledge. This method is vastly used in deep learning for its adaptability and prominent performance [46]. Pre-trained networks are commonly utilized because of their numerous benefits, with one of the most notable being the extensive amount of data they have been trained on which makes them highly efficient and relatively accurate compared to alternative approaches, especially for classification tasks. Such networks could also be subsequently repurposed by others in their respective fields, given the significant effort, time, and accuracy they offer once integrated into new applications [20].

1.3.1 Hyperparameters-Tuning and Fine-Tuning in Transfer Learning:

Hyperparameter tuning and fine tuning are both essential processes in transfer learning, however, both serve distinctive functions in optimizing model’s performance.

Fine-tuning updates the weight of pretrained model on a new dataset, enabling the model to adapt to specific features of the new dataset [47], [48], [49]. This process involves deliberate selection of hyperparameters like batch size, learning rate, and the number of unfrozen layers in order to balance the retention of pre-trained knowledge and adapting to the new dataset [50], [51]. On the contrary, hyperparameter tuning involves systematically exploring and optimizing these hyperparameters for the best performance possible. Methods such as grid search, random search, and Bayesian optimization are commonly employed to methodically explore the hyperparameter space and determine optimal configurations [49], [52]. For instance, a minor learning rate is often used during fine-tuning to prevent overwriting pre-trained weights while adaptive optimization methods such as AdamW or Stochastic Gradient Descent with momentum (SGD momentum) could enhance convergence and stability [53]. The interplay between hyperparameters tuning and fine tuning is specifically essential when the domain shift is substantial between the original dataset and the new dataset since both operations need to work in conjunction to ensure robust generalization [54]. The complimentary nature of these two processes in transfer learning is emphasized by how hyperparameter tuning ensures the training process is efficient and effective and fine tuning adjusting and adapting the model's weights to the new dataset [49], [54].

1.3.2 Overfitting in Transfer Learning:

In transfer learning, overfitting transpires when a pre-trained model, adjusted to a target dataset, conforms too closely to the specific patterns or noise in the target dataset, consequent in suboptimal generalization to unseen data. While transfer learning leverages knowledge acquired previously from a different source domain to enhance performance on specific limited data, the model may overfit during fine-tuning due to insufficient diversity to learn robust features if the data is small or significantly different from the source domain [48], [49], [55]. Overfitting is particularly troublesome in domain medical imaging or natural language processing in transfer learning, where labeled data is not only scarce but could potentially be costly to acquire [56]. To subdue overfitting, researchers have come up with methods such as freeing earlier layers of the pre-trained model to preserve generalizable features while fine-tuning only the last layers and employing other regularization methods such as weight decay and dropout to restrict the model's capacity. In addition, domain adaptation and data augmentation approaches, including adversarial training, have been developed to narrow the disparity between source and target domains which in turn mitigate the chances of overfitting. Regardless of these attempts, overfitting persists as a notable obstacle in transfer learning necessitating cautious tuning of hyperparameters and validation strategies to guarantee robust performance [47], [54], [57], [58], [59].

1.3.3 Underfitting in Transfer Learning:

Underfitting arises when a pre-trained model does not adequately learn the relationship in the dataset which results in subpar performance on both training and validation dataset. It often occurs when the model is too simple or when the fine-tuning process fails to encapsulate the complexities of the target dataset. it frequently occurs in cases where the original dataset and new dataset are substantially different, impairing the model's ability to transfer pertinent knowledge effectively [48], [54], [55]. In medical imaging for instance, a model pre-trained on natural images may find it difficult to generalize to medical data due to shifts between domains. Furthermore, underfitting could emerge due to excessively aggressive regularization, insufficient training epochs or freezing an excessive number of layers during fine-tuning which limits the model's capacity to learn task specific features. To prevail over underfitting, strategies such as unfreezing and fine-tuning additional layers of the pre-trained model, increasing the complexity of the model architecture or utilizing domain adaptation technique to align the source and target domain. Not to mention, providing ample training data and meticulously tuning hyperparameters can enhance the model's adaptability to the new dataset. Resolving underfitting in transfer learning requires a balance between utilizing pre-trained knowledge and enabling the model to learn new task specific patterns [54], [57], [59].

1.3.4 Overfitting Leaks and Verification Processes:

Overfitting leaks pose an essential challenge in machine learning applications as they occur when a model inadvertently learns unrelated patterns or artifacts from the training data where it fails to generalize to the unseen data. This matter often emerges from flawed experimental designs such as data leaks, where information from the validation or testing sets accidentally affects the training process [60], [61]. As an illustration, data leakage may transpire in the preprocessing steps such as normalization or feature selection are applied to the whole dataset before it's divided into training and testing sets, causing the model to peek at the testing data [62]. In vital and crucial sectors like healthcare and finance, overfitting leaks are especially problematic where models trained on leaked data may yield deceptively high-performance metrics, resulting in erroneous decision-making [63]. To address such issues, verification methods were developed to overcome such occurrence and to ensure reliability and generalizability of machine learning models. A commonly employed process of verification is cross validation, particularly k-fold cross validation, where the dataset is split into multiple folds, and the model is repetitiously trained and validated on different subsets. This method provides a more reliable evaluation of model' s performance by averaging results through multiple validation sets, therefore mitigating the chances of overfitting to a single validation split as the figure below illustrates the process [6], [60], [64], [65].

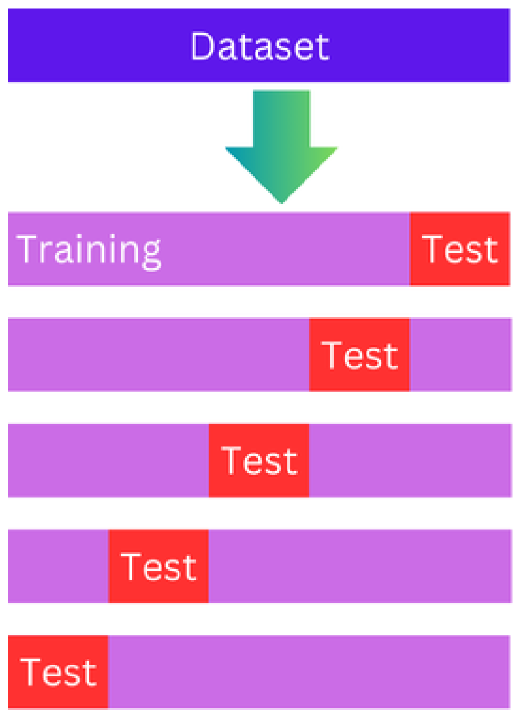

Figure 1.5 K-Fold cross Validation [6], [7].

In hyperparameter tuning, nested cross validation demonstrates to be expressly effective by delineating the hyperparameter optimization process from the final model evaluation. In this approach, an outer loop carries out cross validation for model evaluation, while an inner loop optimizes the hyperparameters on the training folds. Thus, assuring that the validation set used during hyperparameter tuning does sway the final performance metrics [66], [67]. Another indispensable verification method is -

completely held out test set- which only evaluates once after all training and hyperparameter tuning processes have been concluded. This technique offers an unbiased evaluation of the model's performance as the test set remains entirely unobserved during the training process [30]. In fields with inherent or spatial dependencies such as time series or medical data, domain specific splitting strategies are critical. As an instance, in medical imaging, patient-wise splitting ensures that data does not show up in both training and validation sets, consequently, hindering leakage induced by correlated samples [63], [68]. Furthermore, adversarial validation techniques proposed to be a potent approach in detecting and mitigating data leaks by identifying discrepancies between training and test distribution [69], [70]. Recent developments in automated machine learning (AutoML) frameworks such as Optuna and Ray Tune, have presented robust tools for hyperparameter tuning and verification. These tools empower practitioners to automate and streamline data leaks [71]. Additionally, reproducibility frameworks like MLflow and Weight & Biases are progressively adopted to track experimental frameworks, thereby assuring transparency and reducing the chance of overfitting leaks [72]. Through the integration of these verification mechanisms, practitioners can develop models that generalize effectively to the concealed data and generate reliable performance metrics [65].

CHAPTER TWO

IMAGE AUGMENTATION

2.1. Image Augmentation

Image augmentation is part of data augmentation which is referred to as a technique used to increase the dataset size synthetically in efforts to overcome any issues which may rise due to the dataset size such as overfitting or underfitting. It's a process that involves a series of transformations applied to an existing dataset which eliminates the need to use another dataset alongside the existing one. It's an effective and efficient tool to triumph over the necessity to use another dataset. Image augmentation improves models' performance and the ability to recognize patterns and reduces the chance of recognizing the irrelevant pattern. Its benefits do not end there; it also elevates generalization, which provides a wider range of training set in order to aid the model in recognizing unexplored images. In image augmentation, the series of transformations occurrence could be categorized in two categories, first is the Basic Image Augmentation Manipulations, secondly the advanced image augmentation also referred to as Deep Learning Approach as showing in figure 2.1 [8], [73], [74] .

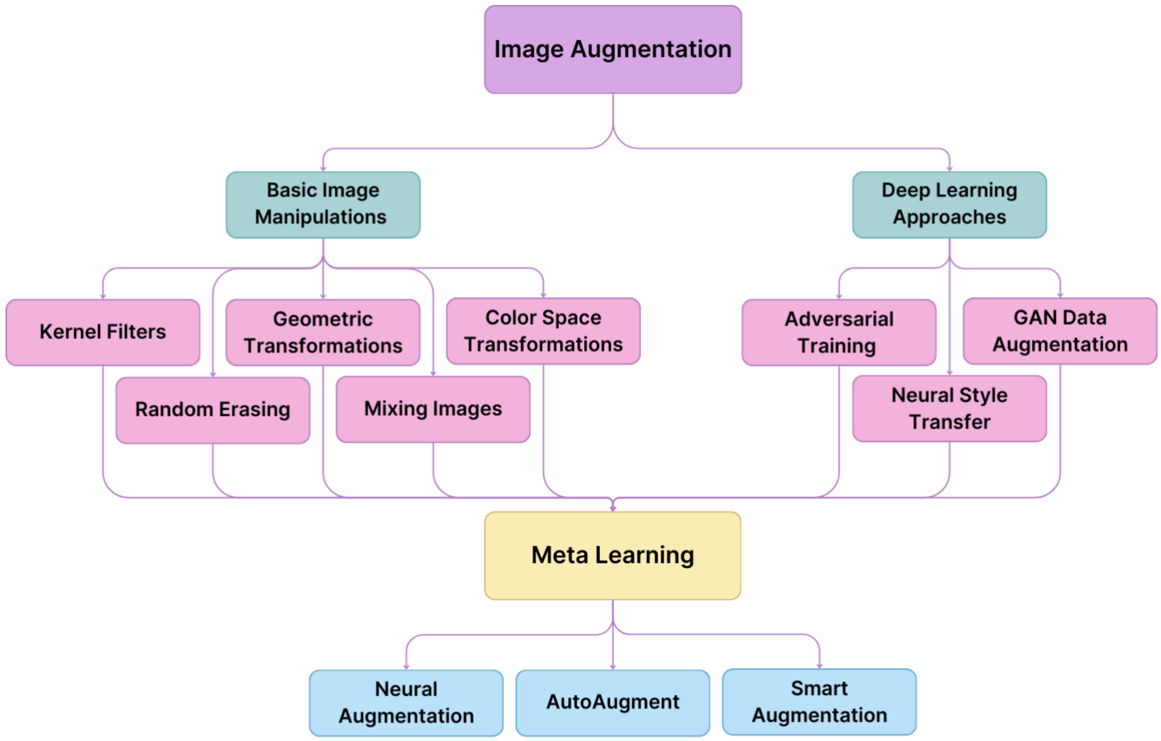

Figure 2.1 Detailed Image Augmentation Taxonomy [8].

2.2. Basic Image Augmentation Manipulation

Basic Image Augmentation Manipulation is a technique that utilized in image augmentation to introduce more pictures with simple and yet effective alteration to an existing dataset, Basic Image Augmentation Manipulation is also referred to as Classical Image Data Augmentation.it consists of five major categories which are Kernel Filtering, Random Erasing, Geometric Transformation, Mixing Images, Color Space Transformations [8], [73], [75].

2.2.1 Color Space Transformations:

The color Space Transformations, also known as color preprocessing, is a technique that relies on the presumption that the colors distributions differ between the training and testing datasets, including contrast Images in a digital form are encoded into three channels. These channels are to be represented in pixel values for an individual value of Red, Green, and Blue (RGB). Lighting and shadows are among the challenges when training and testing an image dataset which this technique offers a valuable insight. Color image preprocessing is Infrequently used, even though it could deliver outstanding results, due to the minimal color differences between training and testing datasets [8], [73], [75], [76]. In the figure below, it illustrates an example of various color space transformations on a moderate demented image. The top left image is the original, in the top right is the image with the effect of HSV (hue, saturation, value) transformation, the bottom left is the image with the effect of LAB transformation, and the bottom right is grey scale effect. Now the least variation exhibited between any of these images in this transformation process is the grey image given the nature of the original image. Such transformation is rarely employed on medical image datasets.

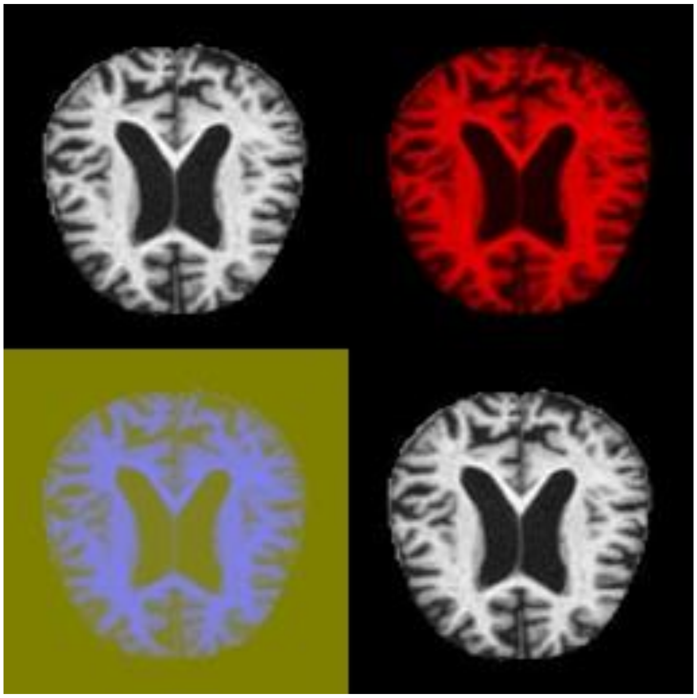

Figure 2.2 Color Space Transformation.

2.2.2. Mixing Images:

Mixing images is a method where fragments from one picture are taken and replaced by another from another picture in the same dataset after these fragments have gone through pixel value averaging from each of RGB channels. While such a technique may not seem an advantageous change to a human viewer, it does show great results in challenge systems and patterns recognition [8], [73], [75]. In the subsequent figure, displays two different and original moderate demented images, unprocessed, which went through the process of image mixing to produce the third image. The first image in the top left is the moderate demented image one, second top right is the moderate demented image two, and thirdly, the bottom center is the product image of mixing both the first and second.

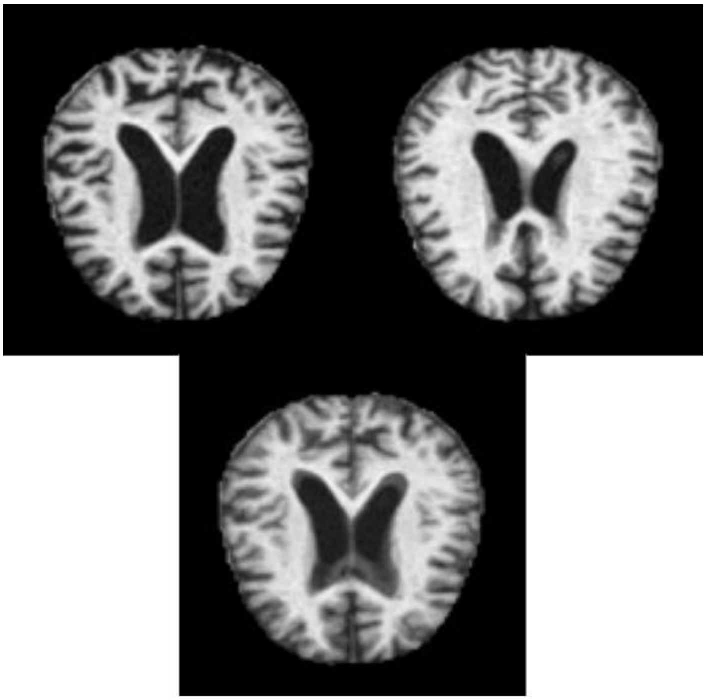

Figure 2.3 Mixed Images.

2.2.3. Geometric Transformation:

In Geometric Transformation, images undergo geometric alteration such as flipping (horizontally or vertically), cropping, and rotation. However, it's worth mentioning that the safety and integrity of an image must be preserved to a certain degree. Some images wouldn't be mistaken for other representations while others might after it has been modified. For example, if an image containing number 6 where to be modified geometrically, it could be misinterpreted as number 9 which it changes the meaning of the picture and deems its safety and integrity to be compromised [8]. Cropping images could very useful not only as an augmentation tool but also as dataset preprocessing tool in preparation to have all images in the dataset sized to model's input size. Also, an image could be cropped from the central portion or any other portion to magnify a picture which is also known an zooming [73], [74]. Another geometric transformation is rotation, where a picture is rotated around its axis between 1 degree to 359 degrees. In rotation, the safety of an image must be kept in mind, as previously mentioned, minimal changes may preserve the label, but in high rotational degrees the safety of a picture may be compromised [73]. The following figure presents the same image, moderate demented, undergo through different geometric transformations, top left shows the original image untampered with, top right is rotated by 45 degrees, bottom left is downscaled by 0.5, and the bottom right is flipped horizontally.

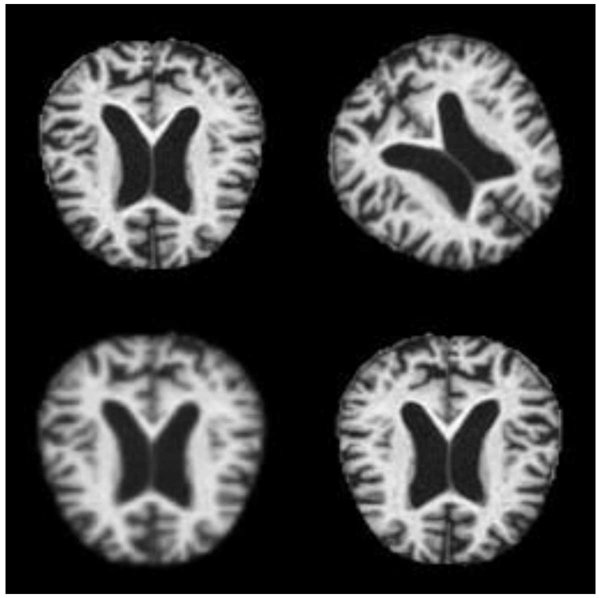

Figure 2.4 Geometric Transformation

2.2.4. Random Erasing:

One of the basic image augmentation manipulations is Random Erasing. It's a method where a random small partition of an image is deleted, usually a square or rectangular shape. The deleted portion could either stay deleted or be filled with random values to further challenge the system. This method pushes the system to learn more features and patterns from the image which could help the model to not focus only on certain features but the entirety of the picture. The method of random erasing, while not considered a part of the geometric transformation, it can be applied in conjunction with any of the basic image augmentation techniques, and so is every method in basic image augmentation techniques [8], [73]. The figure below shows an example of a moderate demented image that went through a random erasing process with the following parameters, minimum area ratio of the erased to the original is 0.02, maximum area ratio of the erased area 0.33 minimum aspect of ratio is 0.3, and maximum aspect ratio is 3.3.

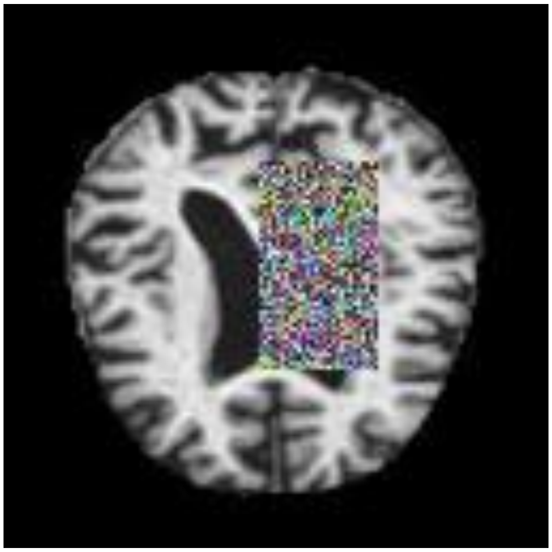

Figure 2.5 Random Erasing.

2.2.5. Kernel Filtering:

Kernel Filtering is a widely known and employed technique in image preprocessing to either produce blurred or sharpened images. The functionality of Kernel filters occurs by convolving an n x n matrix over an image by utilizing a Gaussian filter to make it blurrier or high or high contrast vertical or horizontal edge filter yielding in enhanced images sharpness along the edges. While kernel filters are not limited to using blur filters and sharpening filters, they are some of the traditional methods in kernel filtering. PatchShuffle Regularization is a distinctive kernel filter that has a different approach than the traditional kernel filters which haphazardly interchange between the pixel values within an n x n sliding window. Kernel filters remain comparatively uncharted domain for Data Augmentation. One of the impediments of kernel filters is its similarity to intrinsic mechanisms of Convolutional Neural Networks (CNN). In CNN, the parametric kernels determine the most pragmatic approach to represent images layer by layer [8]. On the contrary to employing a sigmoid activation, which constrains pixel values to a range between 0 and 1. This objective may be accomplished by altering the standard convolutional layer parameters. the adjustment of padding parameters would ascertain the preservation of spatial resolution, and the sub sequent activation layer would sustain pixel values within the interval of 0 to 255. Thus, kernel filters demonstrate superior performance when integrated as a network layer rather than being incorporated into the dataset through data augmentation techniques [8], [77]. The figure below shows the effect of kernel filtering after it was applied to a moderate demented image, the sharpening intensity was set to 5, edge detection value was set to 8, and blurring was set to 5x5 matrix where each element is 1/25.

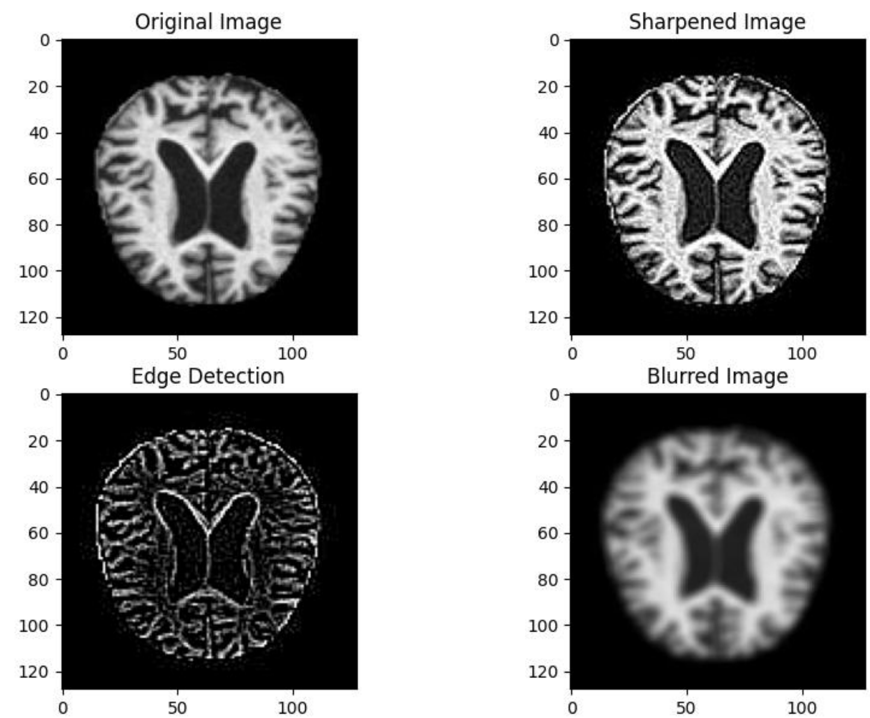

Figure 2.6 Kernel Filtering.

2.3. Deep Learning Approaches

The previously mentioned techniques of image augmentation are implanted at the input stage. Neural networks show a significant aptitude in transforming high-dimensional input data or picture into more concise, lower-dimensional representations with high efficiency. Neural networks demonstrate the ability to convert images into binary classes or n x 1 vectors in flattened layers [8]. Through deliberate manipulation of the sequential processing in neural networks, intermediate representations can be isolated from the network overall structure. Image data is transformed into lower dimensional depiction in fully connected layers which form the feature space. Feature space augmentation occurs within these fully connected layers as opposed to input layer which facilitates the implementation of vector-based augmentation techniques offering new opportunities to further ameliorate model’s robustness and performance. One of the popular methods to address class imbalance is SMOTE, Synthetic Minority Oversampling Technique, which generates new instances in the feature space by adding connecting k-nearest neighbors. Prevalent feature space techniques include adding noise, interpolation, and extrapolation. Autoencoders are peculiarly effective at such purpose, using an encoder to map images into low dimensional vectors and decoder to reconstruct them. Furthermore, feature space could be accomplished by vector isolation representation from CNN, truncating the output layer to produce low dimensional vectors instead of class labels. Such vectors have the ability to train models like Naïve Bayes, SVMs or multiplayer networks. Vector data interpretation present a significant challenge in feature space augmentation. Despite autoencoders can reconstruct vector into images, it requires duplication the CNN’s encoder which in turns lead to more complexity that is time consuming to train [8], [74].

2.3.1. Adversarial Training:

Adversarial training is an approach that employs two or more networks with opposing objectives incorporated within their loss functions. It could serve in two functionalities effectively. First, it can operate as a search algorithm for identifying effective augmentation strategies. Secondly, it addresses the issue of adversarial attacks by providing a mechanism to counteract this circumstance. Adversarial attacks entail a rival network that is trained to generate image augmentations so it leads to misclassification in the classification network. Such attacks, generally restrained to noise injections, have proven to be remarkably effective, challenging conjectures about how models represent images. The success of said attacks underlines the susceptibility of deep learning models to minor yet meticulously devised modification. While Adversarial attacks can be targeted or untargeted, whether the aim is general errors or misclassification, such attacks can reveal weak decision boundaries in models, providing insights beyond standard classification metrics. While Adversarial attacks could be utilized as an excellent evaluation tool, adversarial training could also be used for searching for augmentation. It could be trained to generate augmentation that cause misclassification by restricting the range of augmentation and distortion accessible to an adversarial network, thereby creating an efficient search algorithm [8], [74], [78], [79]. Another method focuses on adversarial networks that modify training labels. For instance, approaches such as DisturbLabel add noise to the loss layer by randomly swapping labels during training, instead of introducing such alteration at the input or hidden layers, indicating that adversarial training also be applied to the label space. Adversarial training has presented auspicious potential in improving robustness against adversarial attacks, however, its ability to mitigate overfitting remains under investigation [8], [74], [78], [79].

2.3.2 GAN-based Data Augmentation:

Generative modeling is one of the techniques widely used in data augmentation and has achieved significant success. generative modeling as the name suggests involves creating artificial data instances that closely resemble the original data while preserving the core characteristics. Amid generative modeling techniques, Generative Adversarial Networks (GANs) have emerged as a favored and powerful method by leveraging adversarial training to accomplish astonishing results in computational speed and output results. Generative Adversarial Networks (GANs) utilize a generator to create synthetic data and a discriminator to assess its authenticity while both networks continuously improve through their interaction. The figure 2.7 below shows the general concept of GANs' structure which can be modified to suit various strategies and approaches [8], [9].

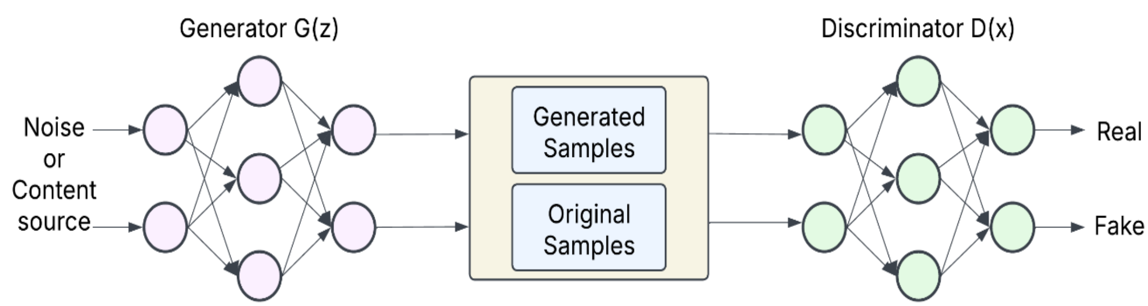

Figure 2.7 Simplified GANs structure [9].

The remarkable abilities of GANs have driven considerable interest in their application for data augmentation. Generative Adversarial Networks (GANs) can elevate the performance of classification models and take on challenges such as limited dataset and class imbalance through its ability to create high quality synthetic data.

2.3.3 Neural Style Transfer:

Neural Style Transfer (NST) is one of the charismatic tactics used in deep learning that modifies the representation of picture within the convolution neural networks. Essentially, the principle is to pick a style from one picture and content from another and merge them to create the output image. While this approach is known for its artistic application, it still serves as a potent tool for data augmentation. It is accomplished through the process of conducting operations on the gram matrix which effectually embodies the stylistic features of an image. The algorithm can create novel pictures that seamlessly combine the structural content of one image with the artistic style of another by fine-tuning the weight assigned to the content and style loss functions. A notable development in this domain is the emergence of Fast Style Transfer which displace the original pre-pixel loss function with a perceptual loss function. This perceptual loss is computed through a pre-trained network which facilitates a quicker and more efficient style transfer. Moreover, the use of instance normalization instead of batch normalization has shown in considerable occasions to substantially improve speed and of quality of stylization. In data augmentation, Neural Style Transfer could be considered as an extension of color space transformation. It doesn't only offer for variation in lighting but also enable the encoding of different textures and artistic styles, providing a variety for stylistic options for generating training images. Style selection for an application can present a challenge. While it is intuitive in applications like self-driving cars (e.g., day to night or summer to winter), it may not be as obvious in other cases. While NST has its disadvantages including the necessity of choosing and applying styles, potential biases stemming from limited sets, and the computational inefficiencies of the original algorithm, which render it unsuitable for large scale tasks [8], [9], [80].

CHAPTER THREE

RELATED WORK

Recent years have seen a considerable and extensive amount of research focused on Alzheimer's disease (AD), particularly in the areas of early diagnosis, detection, and classification. Much of this work has revolved around developing techniques and approaches aimed at achieving optimal performance and practicality for day-to-day use in clinical settings. Various methods have been employed, including the use of different imaging formats, transfer learning with modified or additional layer, and the incorporation of larger datasets for training. These research endeavors have proven highly beneficial in addressing the challenges posed by Alzheimer's disease. A significant collection of recent studies utilizing artificial intelligence and transfer learning has been reviewed and concisely presented in the table below, Table 3.1 [1].

Table 3.1 Summary of all recent related work [1].

References	Computational Techniques	Study Objectives	Datasets	Results	Years
Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks[81]	A deep convolutional neural network is proposed for early-stage Alzheimer's disease diagnosis using brain MRI data analysis, outperforming existing binary classification methods, and outperforming baselines in experiments.	The study offers a deep CNN capable of identifying and classifying Alzheimer's disease, demonstrating superiority and capability in performance on a small dataset, and efficiently approaching imbalanced data to train and learn.	Data obtained from Open Access Series of Imaging Studies OASIS.	The presented model achieves an accurate rate of 93.18%, with precision of 94%, recall of 93%, and a f1-score of 92%.	2018

Table 3.1—Continued Summary of all recent related work [1].

References	**Computational Techniques**	**Study Objectives**	**Datasets**	**Results**	**Years**
Transfer Learning With Intelligent Training Data Selection for Prediction of Alzheimer's Disease [28]	A transfer learning VGG, pre-processing the dataset includes images while applying image entropy to select the most relevant information.	Lowering the reliance on large data training while applying the layer-wise transfer learning to examine the training size impacts.	The dataset has been acquired by the benchmark dataset for deep learning based on Alzheimer's Disease Neuroimaging Initiative (ADNI).	Results technique has shown a 10–20 times smaller data size improvement in accuracy (4–7%) in classification problems involving AD vs. NC, AD vs. MCI, and MCI vs. NC.	2019
Neuroimaging and Machine Learning for Dementia Diagnosis: Recent Advancements and Future Prospects [82]	A comprehensive survey of automated diagnostic methods for dementia that utilizes medical image analysis through machine learning algorithms published in recent years.	To discuss the most recent neuroimaging procedures in the field of dementia diagnosis for clinical applications and Evaluating deep learning approaches in early-stage detection of dementia.	(ADNI), (OASIS), Australian Imaging, Biomarker and Lifestyle Flagship Study of Ageing (AIBL), and CAD Dementia, structural brain MRI scans.	Considering the current diagnostic approaches for AD using MRI scans, it is essential to work on diagnosing other types of dementia, such as FTD, VD, and PD. Deep learning techniques approaches outperform brain images obtained, rather than the conventional ML, in terms of accuracy and early diagnosis of dementia.	2019

Table 3.1—Continued Summary of all recent related work [1].

References	**Computational Techniques**	**Study Objectives**	**Datasets**	**Results**	**Years**
Ensembles of Patch-Based Classifiers for Diagnosis of Alzheimer's Diseases [83]	Feature extractors and a softmax cross-entropy classifier (CNNs), with three distinct models forming the decision-making framework.	Accuracy, overfitting challenges, and key brain landmarks in the left and right hippocampus for clear AD diagnosis.	National Research Center for Dementia (GARD), Gwangju Alzheimer's and Related Dementia dataset	Achieving 90.05% accuracy compared to the other state-of-the-art models on the same dataset.	2019
A Data Augmentation-Based Framework to Handle Class Imbalance Problem for Alzheimer's Stage Detection [84]	TL using data augmentation for 3D Magnetic Resonance Imaging (MRI)	TL for multiclass AD with pre-trained AlexNet model. Two approaches for the brain and 3D brain MRI view while considering an extensive image augmentation to avoid overfitting issues.	From the publicly available dataset (OASIS).	The presented model's accuracy utilizing a 3D view of the brain MRI is 95.11%, whereas using a one-sided view is 98.41%.	2019
Optimized One vs One Approach in Multiclass Classification for Early Alzheimer's Disease and Mild Cognitive Impairment Diagnosis [85]	A pairwise *t*-test feature selection is employed to estimate selected features onto a Partial-Least-Squares multiclass subspace for one vs one output error correction.	To improve accuracy and reduce dependency on large data trained.	the data from the international challenge for automated prediction of MCI from MRI data to address the multiclass classification problem.	The proposed multiclass classification approach outperformed with 67% accuracy, illustrating the robustness towards small fluctuations.	2020

Table 3.1—Continued Summary of all recent related work [1].

References	Computational Techniques	Study Objectives	Datasets	Results	Years
Spatial-Temporal Dependency Modeling and Network Hub Detection for Functional MRI Analysis via Convolutional-Recurrent Network [86]	An end-to-end deep learning Spatial-Temporal convolutional-recurrent neural Network (STNet) model	To predict Alzheimer's disease automatically considering the use of progression and network hub detection implying rs-fMRI time series.	The rs-fMRI time-series data collected from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database1 were studied in this paper.	Experiments on 563 rs-fMRI images from the ADNI database demonstrate improved classification performance over advanced methods and shed light on the pathogenic cascade of AD.	2020
Resting-State fMRI and Improved Deep Learning Algorithm for Earlier Detection of Alzheimer’s Disease [87]	Using an Improved Deep Learning Algorithm (IDLA) which utilizes resting state fMRI along with important un-identifying information such as age and sex. It also utilizes autoencoder customization for the categorization of MCIs vs NCs.	Detecting Early-stage Alzheimer’s disease considers deep neural network data.	ADNI-2 fMRI data in the ADNI database	The suggested approach enhances diagnostic accuracy by approximately 25% over traditional methods. This establishes that incorporating advanced deep learning with brain analysis is a very effective strategy for the early detection of neurological disorders.	2020

Table 3.1—Continued Summary of all recent related work [1].

References	Computational Techniques	Study Objectives	Datasets	Results	Years
Diagnosing Alzheimer's Disease Based on Multiclass MRI Scans using Transfer Learning Techniques [88]	They used whole slide 2-dimension (2D) images to classify AD mild cognitive impairment and normal control subjects using state-of-the-art CNN base models. Also, they evaluated their effectiveness using an AD Neuroimaging Initiative dataset and demonstrated uniqueness using MR images.	Early Alzheimer's diagnosis and classification are crucial for preventing dementia progression in medical image analysis. Therefore, the primary objective was to use Deep learning techniques to detect early Alzheimer's disease.	The study uses the (ADNI) brain MR images.	Three models split data to 70:30 ration training and testing, respectively. The top result was shown by ResNet-101, with 98.37% accuracy and outstanding performance in multiclass classification, is the best out of the three.	2020
Deep Learning of Static and Dynamic Brain Functional Networks for Early MCI Detection [89]	CNN framework to simultaneously learn embedded features from BFNs for brain disease diagnosis.	To learn deeply embedded spatial patterns of the static and dynamic BFNs for eMCI diagnosis.	A rigorously collected, publicly accessible, multisite Alzheimer's Disease Neuroimaging Initiative 2 (ADNI2) dataset	Significant diagnostic performance improvement by almost 10%, including deep learning's efficiency in the preclinical diagnosis of Alzheimer's disease, according to the intricate and multidimensional voxel-wise spatiotemporal patterns of the brain's functional connectomics at rest.	2020

Table 3.1—Continued Summary of all recent related work [1].

References	Computational Techniques	Study Objectives	Datasets	Results	Years
A 3D densely connected convolution neural network with connection-wise attention mechanism for Alzheimer's disease classification [90]	It presents CAM-CNN, a densely connected network with a connection-wise attention mechanism. From pre-processed images, it extracts multi-scale features and integrates layer connections using this attention technique. Tested on baseline MRIs of 968 participants, it distinguishes AD, MCI converters, and non-converters using data from each 3D convolution layer.	The study proposes a deep learning method for efficient detection and prediction of Alzheimer's disease (AD) using a densely connected convolution neural network and connection-wise attention mechanism.	Collected from ADNI.	The proposed algorithm outperformed previous methods in distinguishing AD patients from healthy controls, with a 97.35% accuracy rate, MCI converters vs. Healthy subjects with 87.82%, MCI converters vs. non-converters with 78.79%.	2021
Deep learning-based pipelines for Alzheimer's disease diagnosis: A comparative study and a novel deep-ensemble method [29]	The study uses a deep learning approach to apply transfer learning techniques to CNN architectures pre-trained on Imagenet. The top three networks are AlexNet, Inception-ResNet-v2, ResNet-50, ResNet-101, and GoogLeNet. After fine-tuning, they are combined and classified using an ensemble bagged trees model.	This study presents an automated deep-ensemble approach for dementia-level classification from brain images, compares deep learning architectures, and evaluates robustness in detecting Alzheimer's disease and different dementia levels.	Data acquired from OASIS, KAGGLE, and ADNI.	The proposed strategy, tested on three MRI and one fMRI datasets, achieved an accuracy of 98.51% in binary classification (different levels of dementia recognition) and 98.67% in multiclass classification, surpassing cutting-edge methods.	2021

Table 3.1—Continued Summary of all recent related work [1].

References	Computational Techniques	Study Objectives	Datasets	Results	Years
Brain Asymmetry Detection and Machine Learning Classification for Diagnosis of Early Dementia [91]	The study introduces a new method for early dementia diagnosis using an asymmetry segmentation algorithm, allowing visualization of differences between brain hemispheres, simplifying feature engineering, and offering an advantage over existing methods.	The proposed pipeline offers a cost-effective solution for classifying dementia and possibly other brain degenerative disorders influenced by changes in brain asymmetries.	Acquired from the ADNI database	C-SVM and Q-SVM performed best among SVM variants, with C-SVM achieving 92.5%, 93.0%, 93.0%, and 85.0% accuracy across EMCI vs. NC, AD vs. NC, and AD vs. EMCI. Q-SVM reached 92.5% accuracy in sensitivity and specificity for EMCI vs. NC, AD vs. NC, and AD vs. EMCI. CNN predictions were comparable to other classifiers	2021
Differentiating Dementia with Lewy Bodies and Alzheimer's Disease by Deep Learning to Structural MRI [92]	ResNet was implemented due to its unique characteristics in preserving features in 3D images while performing similarly to other Convolutional Neural Networks.	This study investigates deep learning's ability to differentiate Alzheimer's Disease (AD) from Dementia with Lewy Bodies (DLB) using structural MRI, compared to voxel-based morphometry (VBM).	208 participants, 101 DLB, 69 AD, and 38 controls, which it was obtained from Wellcome Department of Imaging Neuroscience, University College London, UK, www.fil.ion.	Conventional analysis revealed no significant atrophy, while deep learning distinguished DLB from AD with 79.15% accuracy, outperforming the conventional method's 68.41% and highlighting subtle differences it may miss.	2021

Table 3.1—Continued Summary of all recent related work [1].

References	**Computational Techniques**	**Study Objectives**	**Datasets**	**Results**	**Years**
Comparison Of Machine Learning approaches for enhancing Alzheimer's disease classification [93]	The study developed three machine learning-based MRI data classifiers to predict Alzheimer's disease (AD) and infer brain regions. They were compared to each other, SVM, VGGNet, and ResNet, and a transfer learning strategy was applied to improve performance and efficiency.	This study compares three models, one SVM-based and two deep learning algorithms, 3D-VGGNet and 3D-ResNet, for predicting Alzheimer's Disease (AD) and identifying brain regions contributing to disease progression.	A total of 560 images were acquired from ADNI; T1-weighted MR.	The ResNet model outperformed the other two classifiers in detecting Alzheimer's disease (AD) in elderly control subjects with an accuracy of 90% for SVM, 95% for VGGNet, and 95% for ResNet.	2021
Analysis of Features of Alzheimer's Disease: Detection of Early Stage from Functional Brain Changes in Magnetic Resonance Images Using a Fine-tuned ResNet18 Network [94]	proposed fine-tuning model uses ResNet-18, consisting of 3 x 3 filters, 1 x 1 filter, and a fully connected layer; last layer softmax layer. The model adapts pre-trained parameters to the new dataset by unfreezing all layers.	The paper shows a deep learning-based method for predicting MCI, early MCI, late MCI, and AD using hippocampal fMRI data from the ADNI database for early diagnosis.	Data from The ADNI; fMRI dataset. A total of 138 subjects were used for evaluation.	The proposed model performed exceptionally compared to other models with classification accuracy of 99 99.99%, 99.95%, and 99.95% on EMCI vs. AD, LMCI vs. AD and MCI vs. EMCI classification scenarios, respectively.	2021

Table 3.1—Continued Summary of all recent related work [1].

References	Computational Techniques	Study Objectives	Datasets	Results	Years
An Intelligent System for Early Recognition of Alzheimer's Disease Using Neuroimaging [95]	Testing the AD multiclass classification's performance with ResNet18 and DenseNet201.	The study explores the challenge of using randomized concatenated deep features from two pre-trained models that extract discriminative features from MRI images of brain functional networks.	The data was acquired from the ADNI. A total of 138 MRI scans, 25 AD, 25 CN, 25 SMC, 25 EMCI, 13 MCI and 25 LMCI scans.	The proposed model achieved 98.86% accuracy, 98.94% precision, and 98.89% recall in multiclass classification, demonstrating its potential for predicting neurodegenerative brain diseases like Alzheimer's disease.	2022

CHAPTER FOUR

METHODOLOGY

Firstly, the data were acquired from Kaggle, which consists of combined MRI images from multiple sources such as the Alzheimer's Disease Neuroimaging Initiative (ADNI), IEEE, Data.gov, and Cordis EU [96]. The data consisted of a large number of images in various stages that were found to be helpful to this study for generating a robust and accurate system. The dataset was made of 6400 MRI images and consisted of four classes: moderate demented, mild demented, very mild demented, and non-demented. The moderate demented had 64 images, the mild demented class had 896 images, the very mild demented had 2240 images, and the non-demented had 3200 images, as shown in Table 4.2. All of these MRI images were later pre-processed accordingly for training [1], [88].

Table 4.2 Alzheimer's dataset imbalanced [1].

Label	Count
Moderate Demented	64
Mild Demented	896
Very Mild Demented	2240
Non-Demented	3200

Part of the image pre-processing was to ensure we had balanced data, that is, the same number of images per class, to ensure the accuracy of the results. The data augmentation technique is one of many to evade overfitting or underfitting. The minority and the majority classes are another system of representation to which class is referred to that has the lowest number of data and the highest number of data; the minority class here would be the moderate demented class with 64 images followed by mild demented with 896 images, and very mild demented with 2240 images, respectively. The majority class would be the non-demented with 3200 images. Image augmentation was implemented to resolve the data imbalance between classes where all classes, except the non-demented, were augmented to reach 3200 images per class, which resulted in an equilibrium between all classes. Part of the imagining pre-processing phase was performed using Python image augmentation [17], [78]. Table 4.2 shows the images' quantity per class before applying data augmentation. The image augmentation process started with configuring Microsoft Visual Studio Code to Python and installing the Python Augmentor library, then importing the dataset into the program. Then, the image augmentation parameters were customized, as illustrated in Table 4.3 [1].

Table 4.3 Image augmentation specification [1].

Images' Class	Pre Augmentation Quantity	Zoom			Flip Top to Bottom	Sample	Post Augmentation Quantity
		Probability	Min Factor	Max Factor			
Mild Demented	896	0.3	0.8	1.5	0.4	2304	3200
Moderate Demented	64	0.3	0.8	1.5	0.4	3136	3200
Non-Demented	3200	0.3	0.8	1.5	0.4	None	3200
Very Mild Demented	2240	0.3	0.8	1.5	0.4	960	3200

The specifications above in table 4.3 could be further illustrated. It applies a zoom operation randomly with a 30% probability, the zoom factor varies between 0.8 (zoom out) and 1.5 (zoom in) and it applies a vertical flip operation with a 40% probability. Subsequently, it produces as many as the “Sample” indicated of augmented image samples using the mentioned transformation.

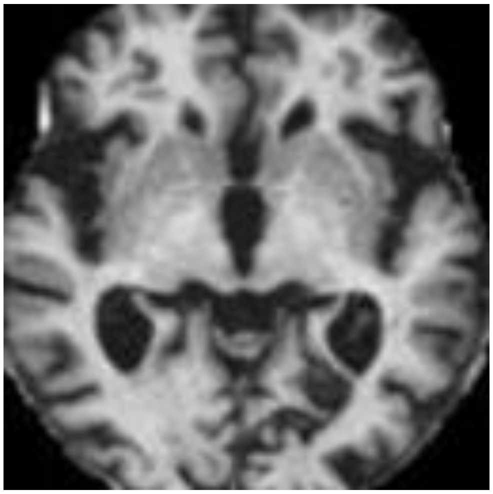

Figure 4.1-A moderated-demented before augmentation

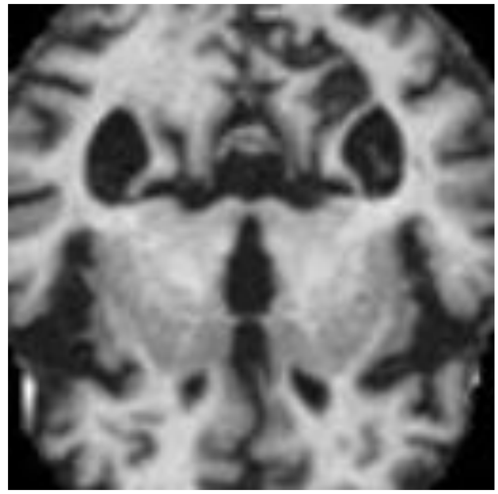

Figure 4.1-B moderated- sample demented sample after augmentation

After augmentation, all data, including non-demented, very mild demented, mild demented, and moderate demented, were 3200 images for each class. Training a deep learning model on an imbalanced dataset could potentially cause overfitting to the model, and to address this issue, this research balanced the dataset using an augmentation technique. Therefore, before applying the data augmentation technique, each class of the data had a significantly different number of images. On the other hand, Table 4.4 lists the number of images for each class after applying the data augmentation technique. Therefore, each class contains the exact number of images [1].

Table 4.4 Alzheimer's dataset balanced [1].

Label	Count
Moderate Demented	3200
Mild Demented	3200
Very Mild Demented	3200
Non-Demented	3200

4.1. Transfer Learning

Transfer learning's foundation lies in the notion that knowledge that has already been learned may be applied effectively and efficiently to address new issues and challenges. Hence, transfer learning necessitates effective and reliable machine learning techniques that retain and recycle previously acquired knowledge [46]. Transfer learning is one of the very well-established methods in deep learning and is mostly used due to its high efficiency, where one would use a pre-existing trained network for their application. This method is vastly preferred due to its efficiency, and, in some cases, it yields the best results with proper fine-tuning. To employ transfer learning, there are two main ways to perform this: training the original network without any modifications or alterations to the layers and fine-tuning where adjustments would be applied to tailor-fit the chosen network for the job. Some examples of pre-existing transfer learning networks are ResNet-50, AlexNet, and VGG19 [1], [97].

4.1.1. AlexNet:

AlexNet has 25 layers and a depth of 8 and is considered among the most compact pre-trained networks for transfer learning. Convolution, normalizing, pooling, and a Rectified Linear Unit (ReLU) are fully linked layers in the order in which AlexNet's layers are arranged. AlexNet is a great example of a simple, small, and yet effective pre-existing network for its size [1], [97], [98], [99]. The network's depth layers comprises of eight layers, five of which are convolutional layers and three fully connected layers. The convolutional layers are responsible for feature extraction, while the fully connected layers perform classification. The first convolutional layer operates on the input image with 96 filters of size 11x11, followed by a max-pooling layer. Mathematically, the output of the first convolutional layer can be represented as [99]:

$$O_1 = f(W_1 * I + b_1) \tag{1}$$

Here I is the input image, W_1 represents the weights of the convolutional filters, b_1 is the bias, $*$ denotes the convolution operation, f is the activation function (ReLU), O_1 is the output feature map. In Subsequent convolutional layers, filters of varying sizes are applied, the number of filters increases as the network deepens. The ReLU activation function, defined as $f(x) = \max(0, x)$, replaced traditional sigmoid and tanh functions. Its non-saturating nature mitigated the vanishing gradient issue, enabling faster training. Max-pooling layers are interspersed between certain convolutional layers to reduce the spatial dimensions of the feature maps, thus decreasing the number of parameters and computational demands. The max-pooling operation can be expressed as [99]:

$$P(x, y) = max_{i,j} A(x + i, j + j) \tag{2}$$

Where A is the input feature map, P is the output pooled feature map, (x, y) are the coordinates of the output, and (i, j) are the offsets within the pooling window.

After the convolutional layers, the feature maps are flattened and fed into three fully connected layers. The output of the fully connected layers can be represented as [99]:

$$O_{fc} = f(W_{fc} \cdot A_{flat} + b_{fc}) \quad (3)$$

Where A_{flat} is the flattened feature map, W_{fc} represents the weights of the fully connected layer, b_{fc} is the bias, f is the activation function (ReLU or softmax) and O_{fc} is the output. AlexNet also employed dropout regularization in the fully connected layers as a precautionary measure to prevent overfitting. During training, dropout randomly sets a fraction of the neuron outputs to zero, forcing the network to learn more generalized features [99]. The network was trained on the ImageNet dataset, a large-scale image classification dataset with millions of images. The sheer scale of the dataset, combined with AlexNet's architecture and training techniques, allowed the network to achieve a significant improvement in classification accuracy [97], [98], [99].

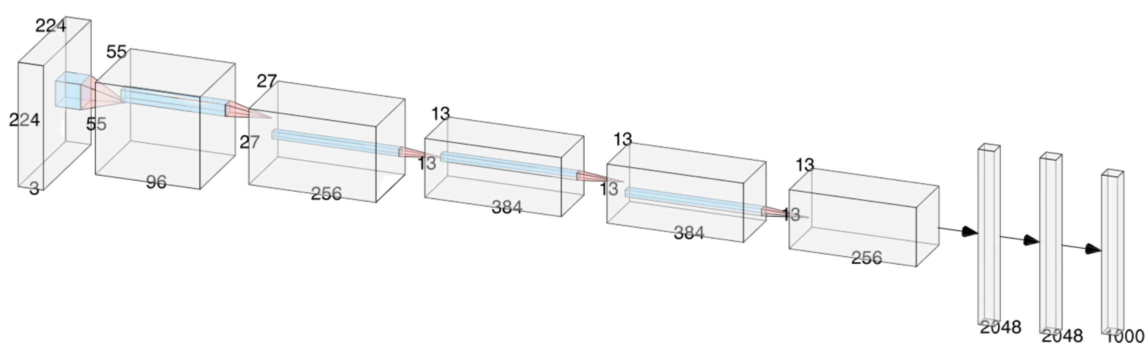

Figure 4.2 AlexNet Network Architecture diagram [10].

4.1.2. ResNet-50:

The residual network has multiple versions in which they are different mainly on how many layers each has, and ResNet-50 has 50 layers consisting of 1 × 1, 3 × 3, 7 × 7 convolution filters: 48 convolution, 1 average pool, and 1 MaxPooling layer. The Resnet-50 architecture propagates the added results of the convolution layer and input called residue. This lowers the overhead associated with the propagation of more characteristics. Thus, the over-fitting issue is diminished to enable quicker optimization of bigger networks [1], [97], [98]. A fundamental component of ResNet-50 is the residual block, which can be expressed mathematically as [11], [100]:

$$H(x) = F(x) + x \tag{4}$$

Where x is the input to the residual block, $F(x)$ represents the residual mapping, typically a series of convolutional layers, and $H(x)$ is the output of the residual block. This equation illustrates that the network learns to approximate the residual function $F(x)$, which is added to the original input x. The "skip connection" or "shortcut connection" facilitates the direct flow of gradient through the network, mitigating the vanishing gradient issue. If optimal mapping is simply the identity function, the network can adapt by setting $F(x) = 0$. The residual blocks in ResNet-50 generally utilizes what is known as a bottleneck architectures, which contain three convolutional layers: A 1x1 convolution to reduce dimensionality, A 3x3 convolution for feature extraction, and A 1x1 convolution to restore dimensionality [100]. The bottleneck design reduces the computational cost of the 3x3 convolutions, which are the most computationally intensive. The number of filters in these convolutional layers varies across the stages, with the filters count doubling at the beginning of each stage. This design enables the network to learn increasingly complex features as it progresses while the residual connection allows the network to adaptively learn an identity function. If additional layers within the residual block are deemed unnecessary, their weights can be adjusted to a zero, and the block will essentially act as an identity mapping. This characteristic is critical for training deep networks [100]. ResNet-50 achieves its depth stacking multiple residual blocks, enabling it to learn highly complex representations. The residual connections ensure efficient gradient propagation, allowing for successful training of very deep models. The final layers of ResNet-50 typically involve global average pooling and a fully connected layer for classification [11], [100].

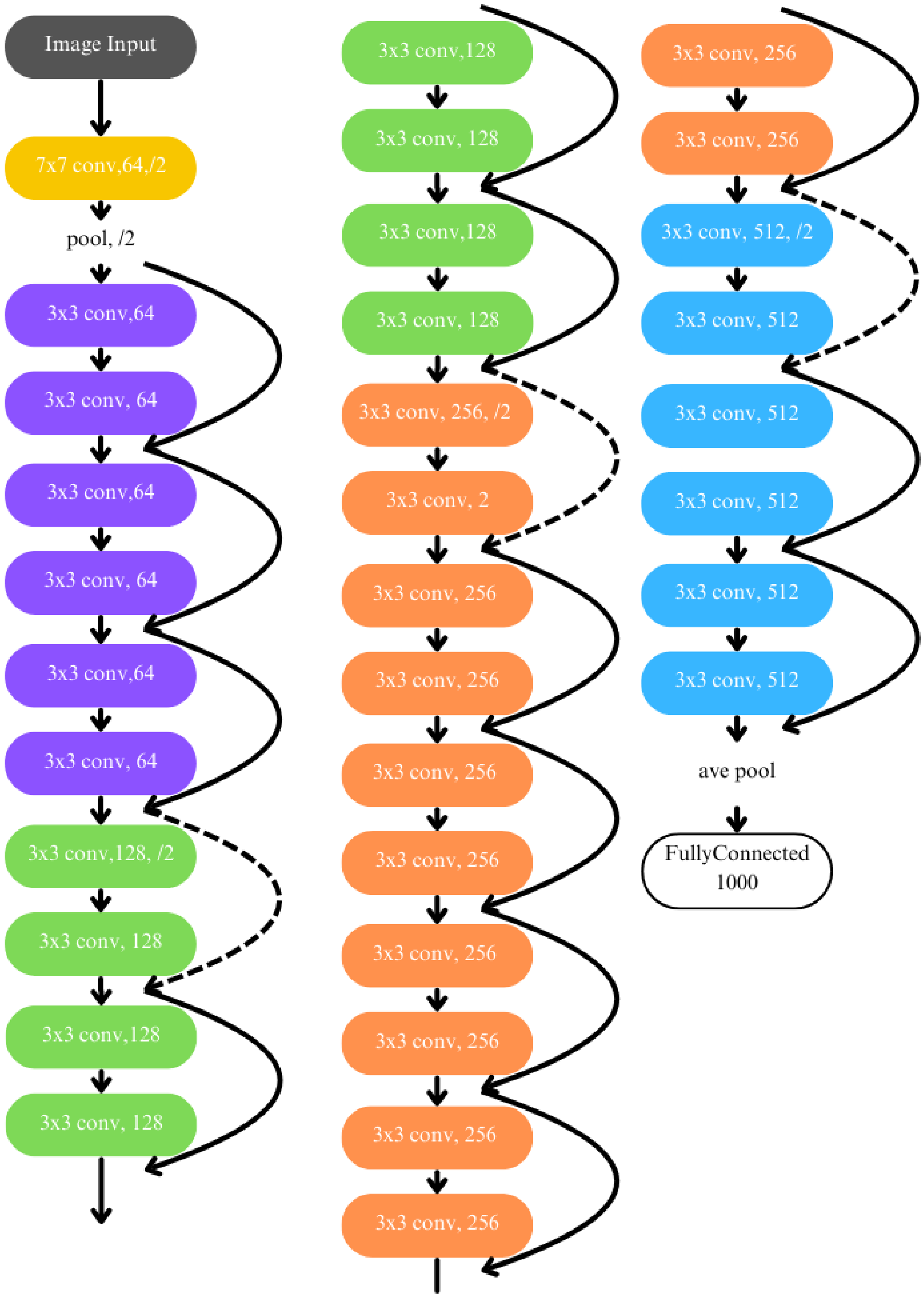

Figure 4.3 Resnet-50 Network Architecture diagram [11].

4.1.3. GoogleNet (InceptionV3):

GoogleNet is another pre-trained network example that consists of 144 layers and works on input images of size 224 × 244 × 3. The network depth layers are 22. The GoogleNet inception is created by combining the output of abundant simultaneous convolution layers with capricious filter sizes. As a result, the concatenated feature map comprises features from the same input attained using several filtering spaces. The auxiliary categorization output layers that result in a deeper network are another trait of GoogleNet [1], [97], [98], [101]. Furthermore, Inception v3 enhances the Inception module by emphasizing factorization and dimensionality reduction to improve efficiency and accuracy[102]. First, it factorizes large 5x5 convolutions into two consecutive 3x3 convolutions, reducing the number of parameters and computational cost while maintaining the same receptive field. Second, it extends this factorization to NxN convolutions by breaking them into 1xN and Nx1 convolutions. For example, a 3x3 convolution is divided into a 1x3 convolution followed by a 3x1 convolution, further reducing computational complexity. Third, it incorporates auxiliary classifiers at intermediate layers, similar to those in Inception v1, to enhance gradient flow during training. Fourth, it introduces label smoothing, a regularization technique that softens hard labels during training [101], [102]. This method helps prevent overfitting and improves the network's generalization ability. The factorization of a 5x5 convolution could be represented mathematically as [102]:

$$O_{5x5} \approx f\left(W_{3x3_2} * f\left(W_{3x3_1} * I + b_{3x3_1}\right) + b_{3x3_2}\right) \quad (5)$$

Here I is the input, W_{3x3_1} and W_{3x3_2} denote the weights of the two sequential 3x3 convolutions, b_{3x3_1} and b_{3x3_2} are the biases, f is the activation function (ReLU), and

O_{5x5} represents the approximate output of the 5x5 convolution. Similarly, The factorization of an NxN convolution into 1xN and Nx1 convolutions can be represented as [102]:

$$O_{NxN} \approx f(W_{Nx1} * f(W_{1xN} * I + b_{1xN}) + b_{Nx1}) \tag{6}$$

Where I is the input, W_{1xN} and W_{Nx1} are the weights of the 1xN and Nx1 convolutions, b_{1xN} and b_{Nx1} are the biases, f is the activation function (ReLU), and O_{NxN} is the approximate output of the NxN convolution. Label smoothing modifies the target probability distribution, typically represented as a one-hot vector, to a smoother distribution. Mathematically, if y is the original one-hot label and ϵ is a small constant, the smoothed label y' can be represented as [102]:

$$y' = (1 - \epsilon)y + \frac{\epsilon}{K} \tag{7}$$

Where K is the number of classes. Inception v3's architectural refinements, particularly factorization and label smoothing, resulted in a significant improvement in image classification accuracy and efficiency compared to its predecessor. The network's depth combined with its optimized Inception modules, enabled it to learn highly complex representations while maintaining a reasonable computational cost [101], [102].

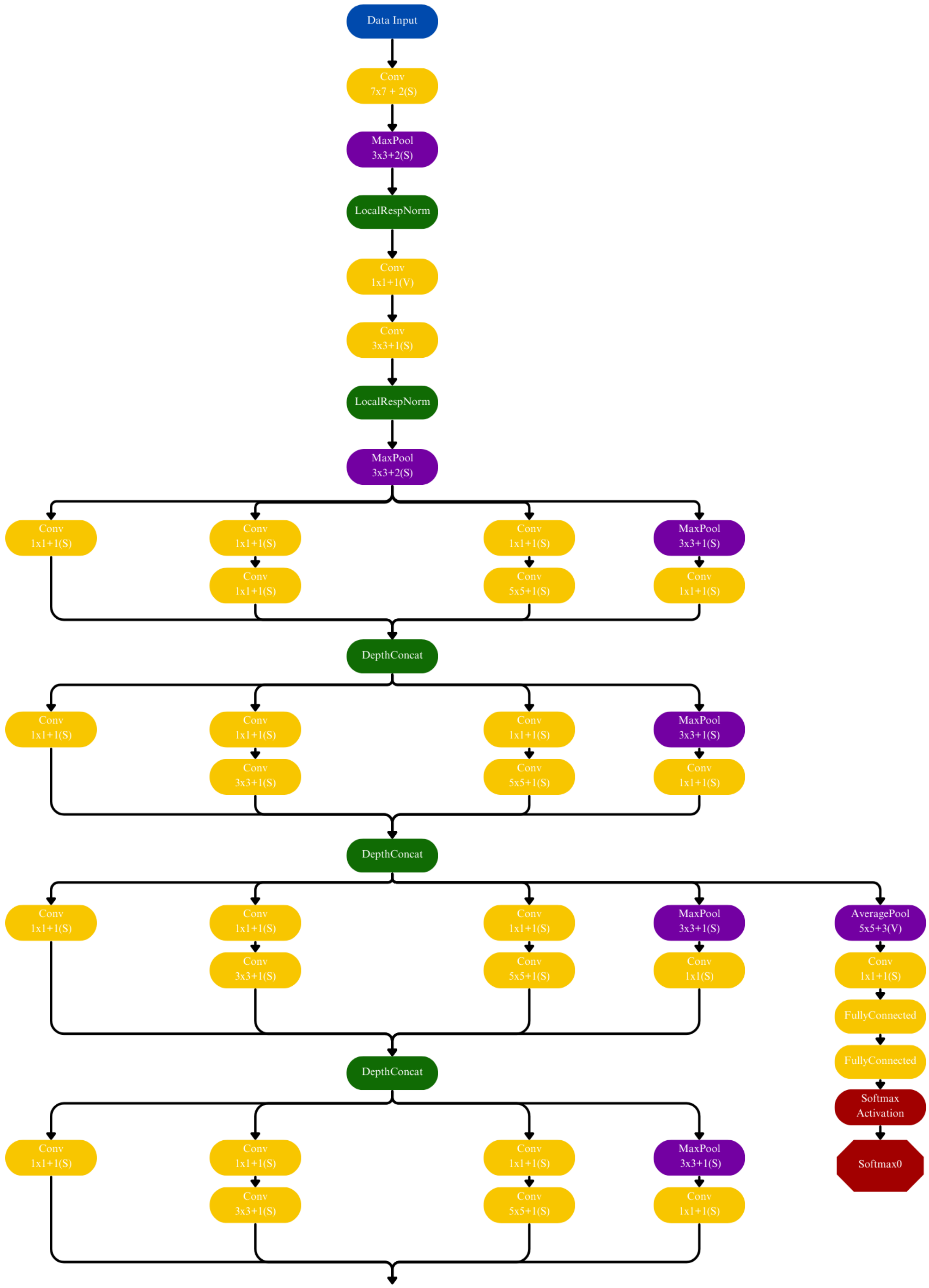

Figure 4.4 GoogLeNet Network Architecture diagram [12].

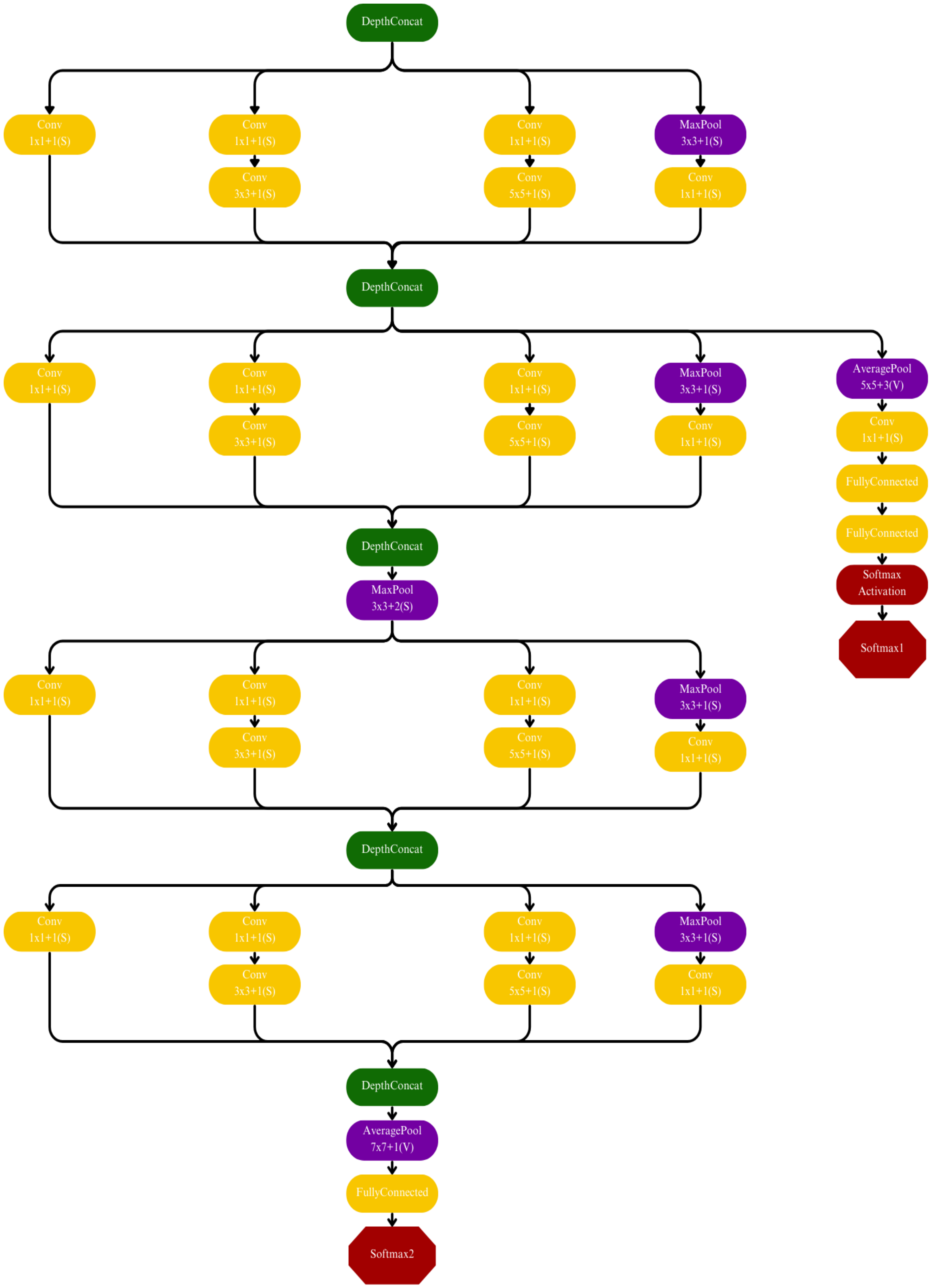

Figure 4.4—continued GoogLeNet Network Architecture diagram [12].

4.1.4. SqueezeNet:

SqueezeNet has 50 times less parameters in comparison with AlexNet and it was trained on over a million photos. The groundwork for this network is laid out as a fire model that comprises the squeeze layer and expand layer. Only 1 × 1 filters are used in the squeeze layer, which feeds into an expand layer that combines 1 × 1 and 3 × 3 convolution filters [1], [98]. SqueezeNet achieves its efficiency through the implementation of the "fire" module which systematically reduce parameters. The main concept is to minimize channel dimensionality before executing the computationally expensive 3x3 convolutions. If C_{in} represents the number of input channels to the fire module and the squeeze layer, a 1x1 convolution, is reduced to S_{1x1} output channels then the parameters in the Squeeze layer (P_{squeeze}) could expressed as the following [103]:

$$P_{squeeze} = C_{in} \times S_{1x1} 1 \times 1 \times 1 = C_{in} \times S_{1x1} \tag{8}$$

Fundamentally, $S_{1x1} \ll C_{in}$, which serves as the basis of SqueezeNet's efficiency. Then, the expand layer combines the outputs of 1x1 and 3x3 convolutions respectively denoted as E_{1x1} and E_{3x3} output channels. The 1x1 expand layers parameters (P_{expand_1x1}) is expressed as [103]:

$$P_{expand_{1x1}} = S_{1x1} \times E_{1x1} \times 1 \times 1 = S_{1x1} \times E_{1x1} \tag{9}$$

And for the 3x3 expand layer (P_{Expand_3x3})[103]:

$$P_{expand_{3x3}} = S_{1x1} \times E_{1x1} \times 3 \times 3 = 9 \times S_{1x1} \times E_{3x3} \tag{10}$$

The sum total parameters in the fire module (P_{fire}) are [103]:

$$\begin{gathered} P_{fire} = P_{squeeze} + P_{expand_{1x1}} + P_{expand_{3x3}} = \\ C_{in} \times S_{1x1} + S_{1x1} \times E_{1x1} + 9 \times S_{1x1} \times E_{3x3} \end{gathered} \tag{11}$$

The key reduction arises from S_{1x1} being considerably smaller than C_{in}, effectively lowering the computational load, especially for the 3x3 convolution, which incurs a multiplicative factor of 9. The total output channels of the expand layer, $E_{total} = E_{1x1} + E_{3x3}$, often exceeds S_{1x1}, ensuring feature expansion following the squeeze. The fundamental principle that SqueezeNet utilizes is [103]:

$$S_{1x1} \ll C_{in} \tag{12}$$

And that the total output of the expand layer is [103]:

$$E_{total} = E_{1x1} + E_{3x3} \tag{13}$$

Where often [103]:

$$S_{1x1} < E_{total} \tag{14}$$

This design substantially reduces the number of parameters while maintaining accuracy, making SqueezeNet a compact and efficient neural network architecture [103].

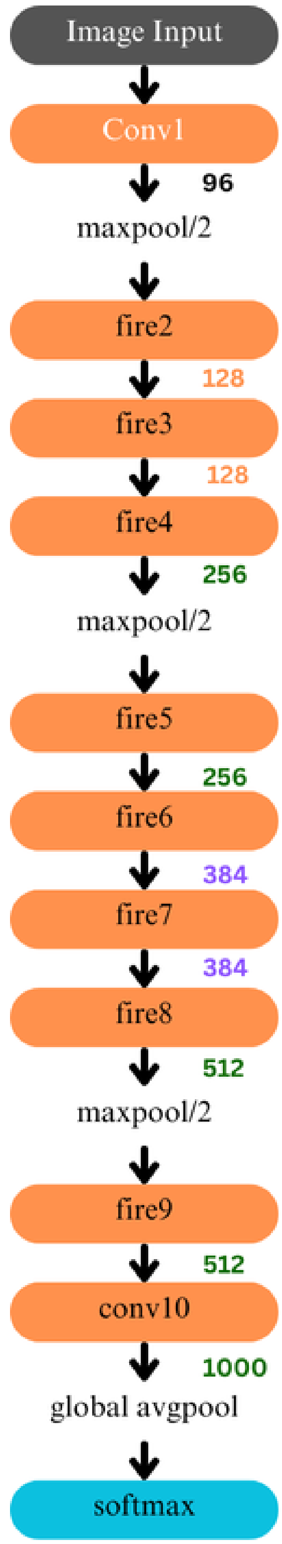

Figure 4.5 SqueezeNet Network Architecture diagram [13].

The diagram below in Figure 4.6 outlines the workflow of the system in detail, such as AlexNet, ResNet-50, and GoogleNet, since they were the main focus of the implantation. Once the dataset has been balanced and augmented, using Python, the remaining process would be carried out in MATLAB.

Figure 4.6 Detailed workflow for the proposed model [1].

After the first phase of image pre-processing was completed, the dataset was called Matlab, where it was stored and organized according to its labels. Then, the data were split randomly into a training set and a testing set, 80% training and 20% testing, for each category. The pre-processing continues to resize all images according to each pre-trained network input layer's specification, such as GoogleNet 224 × 224, and converting from grayscale to RGB, which concluded the pre-processing phase, thus deeming them ready to be injected into the pre-trained network after each pre-trained network has been fine-tuned [1].

CHAPTER FIVE

RESULTS

This section illustrates the experiment setup, result, and evaluation of this research. Pre-trained neural network classifiers, which are AlexNet, ResNet-50, and GoogleNet (InceptionV3), were employed. Each transfer learning model was trained to classify images into four categories: Mild Demented, Moderate Demented, Non-Demented, and Very Mild Demented. Each dataset class contains 3200 images; thus, the total images are 12,800 [1].

Each pre-trained neural network classifier is evaluated using two types of evaluations: (1) k-Fold Cross-Validation and (2) Partitioning the Dataset into Training and Test Sets [1].

5.1. k-Fold Cross-Validation

K-fold cross-validation is a method used for evaluating the performance of machine learning and deep learning models. Specifically, we employed 10-fold cross-validation, in which each dataset is divided into 10 sets; 9 sets are applied for training, and the remaining set is used for testing. The process of training and testing is repeated k times, with k being 10 in our case [104]. We utilized a dataset of 12,800 images, employing the 10-fold cross-validation method for both training and testing purposes [1], [104].

Tables 5.1–5.4 display the overall confusion matrix and the average accuracy of the 10-fold cross-validation. Table 5.1 presents the performance of the AlexNet model based on 10-fold cross-validation. The diagonal values represent the accuracy of correct predictions produced by the AlexNet model for each class of Alzheimer's disease. For example, the class Mild Demented is correctly predicted with 98.66% accuracy. Additionally, Moderate Demented is correctly predicted at 100.00%. Non-Demented and Very Mild Demented are correctly predicted with accuracies of 98.91% and 94.66%, respectively. The average accuracy is 91.63%. The same role is used for Tables 5.2–5.4 but for different pre-trained models. It is clear that the performance of the AlexNet model exceeded the other models while SqueezeNet ranked the lowest among them. The results showed a Mild Demented accuracy of 83.94%, Moderate Demented accuracy of 98.91%, Non-Demented accuracy of 88.13%, Very Mild Demented accuracy of 74.53%, and an overall accuracy of 86.37% [1].

Table 5.1 AlexNet 10-fold cross-validation [1].

	AlexNet Overall Confusion Matrix				**Average Accuracy**
Mild Demented	**98.66%**	0.34%	0.03%	0.97%	98.05%
Moderate Demented	0.00%	**100.00%**	0.00%	0.00%	
Non-Demented	0.03%	0.03%	**98.91%**	1.03%	
Very Mild Demented	4.13%	0.25%	0.97%	**94.66%**	
	Mild Demented	Moderate Demented	Non-Demented	Very Mild Demented	

Table 5.2 InceptionV3 10-fold cross-validation [1].

	InceptionV3 Overall Confusion Matrix				**Average Accuracy**
Mild Demented	**98.59%**	0.06%	0.09%	1.25%	97.80%
Moderate Demented	0.06%	**99.94%**	0.00%	0.00%	
Non-Demented	0.28%	0.00%	**98.44%**	1.28%	
Very Mild Demented	4.25%	0.06%	1.47%	**94.22%**	
	Mild Demented	Moderate Demented	Non-Demented	Very Mild Demented	

Table 5.3 ResNet-50 10-fold cross-validation [1].

	ResNet-50 Overall Confusion Matrix				**Average Accuracy**
Mild Demented	**92.88%**	1.03%	0.63%	5.47%	91.11%
Moderate Demented	2.97%	**95.19%**	0.00%	1.84%	
Non-Demented	2.44%	0.47%	**93.16%**	3.94%	
Very Mild Demented	11.63%	1.25%	3.91%	**83.22%**	
	Mild Demented	Moderate Demented	Non-Demented	Very Mild Demented	

Table 5.4 SqueezNet10-fold cross-validation [1].

	ResNet-50 Overall Confusion Matrix				**Average Accuracy**
Mild Demented	**83.94%**	0.53%	2.19%	13.34%	86.37%
Moderate Demented	0.66%	**98.91%**	0.03%	0.41%	
Non-Demented	1.28%	0.00%	**88.13%**	10.59%	
Very Mild Demented	9.69%	0.22%	15.56%	**74.53%**	
	Mild Demented	Moderate Demented	Non-Demented	Very Mild Demented	

5.2. Partitioning the Dataset into Training and Test Sets

The dataset was split into 80:20 portions, so 80% of the images were used for training, and 20% were used for testing. For the training parameter, the initial learning rate is set to 0.0010, and the maximum number of epochs is set to 50. As for optimization measures, a stochastic gradient descent with momentum (SGDM) was applied. The experiment was performed on a machine with the following hardware: an Intel® Core™ i-7 processor 9th generation and an NVIDIA GeForce RTX 2070 GPU. We utilized MATLAB R2023a [1].

The classification accuracy, precision, recall, and F1 score were calculated to test and evaluate the experiment conducted in this study. The classification accuracy is shown in Equation (15). TP, TN, FP, and FN refer to true positive, true negative, false positive, and false negative, respectively. Precision is computed using Equation (16), recall using Equation (17), and F1-score using Equation (18) [1], [105], [106].

$$\text{Accuarcy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{15}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{16}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{17}$$

$$\text{F1} - \text{Score} = 2 * \frac{\text{Precision} \times \text{Recall}}{(\text{Precision} + \text{Recall})} \tag{18}$$

Table 5.5 and figure 5.1 illustrate the evaluation of this research work, which includes four different pre-trained neural networks. Each pre-network is responsible for performing multi-classification for Alzheimer's MRI Scans. According to Table 5, AlexNet is superior to other pre-trained networks. It is significant to mention that GoogleNet (Inception V3) and ResNet-50 performed comparable performances with similar margins of accuracy (between 92% and 94.776%). SqueezNet shows the lowest performance compared to other pre-trained neural networks across all evaluation metrics as the difference in performance is considerable [1].

Table 5.5 Transfer learning evaluation [1].

Model	Accuracy	Precision	Recall	F1-Score
AlexNet	96.616%	96.621%	96.619%	96.620%
GoogleNet (InceptionV3)	94.776%	94.784%	94.775%	94.779%
ResNet-50	94.363%	94.341%	94.363%	94.352%
SqueezNet	91.602%	92.207%	91.601%	91.904%

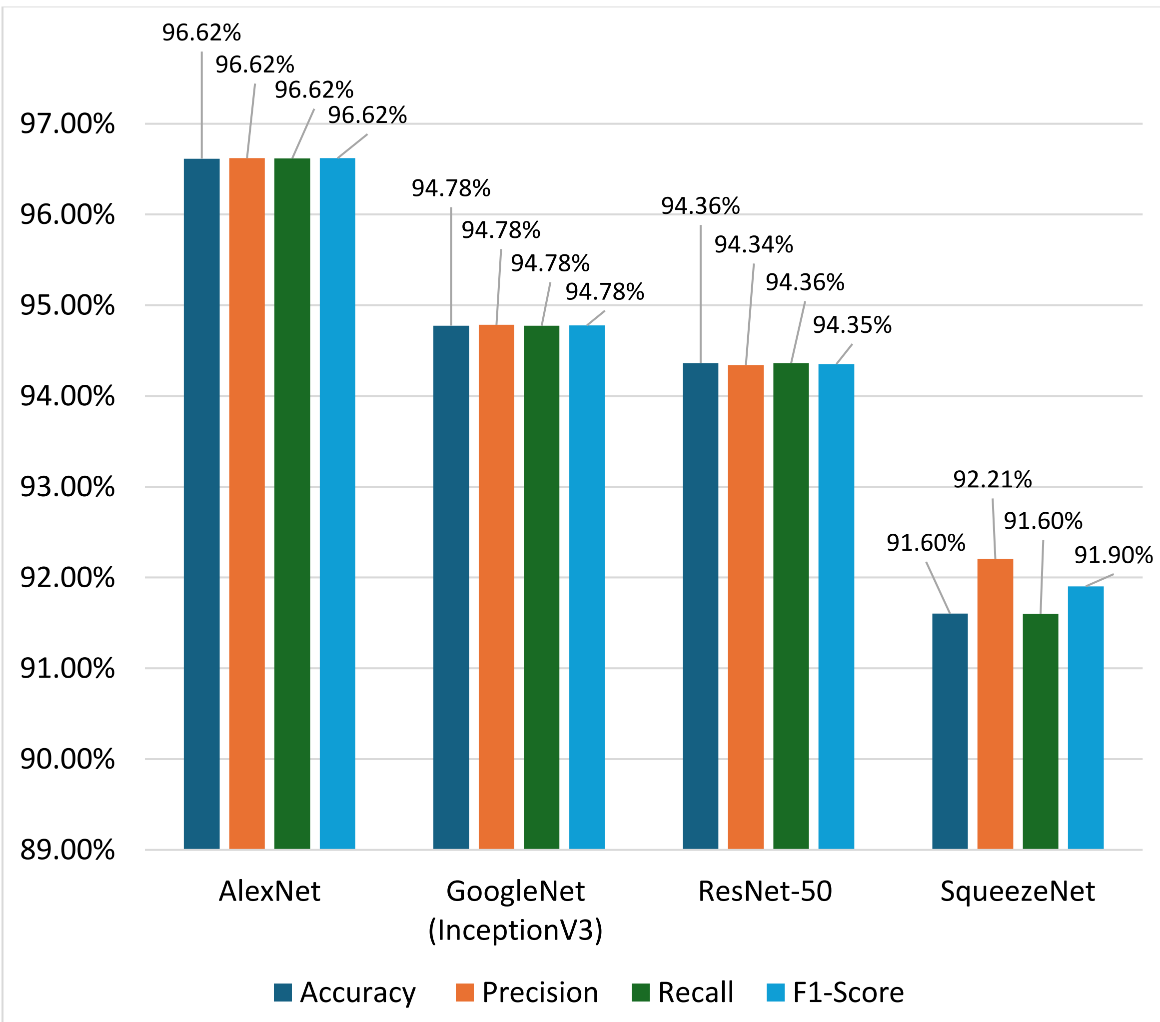

Figure 5.1 Transfer Learning Evaluation Comparison.

As Figure 5.1 illustrates, AlexNet performance was significantly better than the other networks in all evaluation metrics. Also, it is worth mentioning that GoogleNet and ResNet-50 deliver similar results across all metrics both comparatively and independently, maintaining consistency in their metrics. On the other hand, SqueezeNet exhibited some variation across individual metrics—for instance, its Recall was slightly lower compared to its Precision and F1-Score, indicating differences in how well it handles certain aspects of evaluation.

CHAPTER SIX

CONCLUSION AND FUTURE WORK

We conducted Alzheimer’s disease detection and classification into four distinct categories using transfer learning with pre-trained networks, specifically AlexNet, ResNet-50, GoogleNet (InceptionV3), SqueezeNet. The use of transfer learning in deep learning effectively addresses the challenges associated with training networks on limited datasets. The aforementioned models were trained and tested on a total of 12,800 images after they went under image augmentation process. The architectures of these pre-trained networks were modified to accommodate the acquired MRI images, which were categorized into four classes: non-demented, mild demented, moderate demented, and very mild demented. The dataset was sourced from Kaggle and compiled from various repositories, including the Alzheimer’s Disease Neuroimaging Initiative (ADNI), IEEE, Data.gov, and Cordis EU. The performance of each model was assessed with a focus on multi-class classification. Among the models, AlexNet outperformed the others, while GoogleNet (InceptionV3) and ResNet-50 achieved comparable accuracy levels, ranging between 92% and 94.776%. To conclude, while specialized deep learning models for medical imaging detection may produce varying outcomes, transfer learning demonstrates significant promise, particularly when working with limited data. A notable limitation of this study is the relatively small dataset compared to other domains where deep learning is applied even though it’s still considered large in this domain. For future research, exploring the integration of Generative Adversarial Networks (GANs) or Deep Convolutional Generative Adversarial Network (DCGAN) or Self-Organizing Maps

(SOMs) with CNNs or combining the current model with other state-of-the-art models could lead to a more robust system. Our findings highlight the value and potential of deep learning, especially transfer learning, as a valuable tool for advancing medical research in Alzheimer's disease detection using MRI imaging. While these systems currently require medical oversight for implementation, their encouraging outcomes indicate potential as valuable support tools for healthcare professionals in the foreseeable future [1].

REFERENCE

[1] S. Alqahtani, A. Alqahtani, M. A. Zohdy, A. A. Alsulami, and S. Ganesan, "Severity Grading and Early Detection of Alzheimer's Disease through Transfer Learning," *Information*, vol. 14, no. 12, Art. no. 12, Dec. 2023, doi: 10.3390/info14120646.
[2] "2022 Alzheimer's disease facts and figures," *Alzheimers Dement.*, vol. 18, no. 4, pp. 700–789, Apr. 2022, doi: 10.1002/alz.12638.
[3] "Deep Learning." Accessed: Mar. 22, 2024. [Online]. Available: https://www.mathworks.com/discovery/deep-learning.html
[4] "What Is a Convolutional Neural Network? | 3 things you need to know." Accessed: Feb. 01, 2025. [Online]. Available: https://www.mathworks.com/discovery/convolutional-neural-network.html
[5] I. Tabian, H. Fu, and Z. Sharif Khodaei, "A Convolutional Neural Network for Impact Detection and Characterization of Complex Composite Structures," *Sensors*, vol. 19, no. 22, Art. no. 22, Jan. 2019, doi: 10.3390/s19224933.
[6] J. Kaliappan, A. R. Bagepalli, S. Almal, R. Mishra, Y.-C. Hu, and K. Srinivasan, "Impact of Cross-Validation on Machine Learning Models for Early Detection of Intrauterine Fetal Demise," *Diagnostics*, vol. 13, no. 10, Art. no. 10, Jan. 2023, doi: 10.3390/diagnostics13101692.
[7] V. H. Phung and E. J. Rhee, "A High-Accuracy Model Average Ensemble of Convolutional Neural Networks for Classification of Cloud Image Patches on Small Datasets," *Appl. Sci.*, vol. 9, no. 21, Art. no. 21, Jan. 2019, doi: 10.3390/app9214500.
[8] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, p. 60, Dec. 2019, doi: 10.1186/s40537-019-0197-0.
[9] A. Mikolajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," in *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, Swinoujście: IEEE, May 2018, pp. 117–122. doi: 10.1109/IIPHDW.2018.8388338.
[10] "NN SVG." Accessed: Mar. 13, 2025. [Online]. Available: https://alexlenail.me/NN-SVG/AlexNet.html
[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," Dec. 10, 2015, *arXiv*: arXiv:1512.03385. doi: 10.48550/arXiv.1512.03385.
[12] C. Szegedy *et al.*, "Going Deeper with Convolutions," Sep. 17, 2014, *arXiv*: arXiv:1409.4842. doi: 10.48550/arXiv.1409.4842.
[13] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," Nov. 04, 2016, *arXiv*: arXiv:1602.07360. doi: 10.48550/arXiv.1602.07360.
[14] A. Mehmood *et al.*, "A Transfer Learning Approach for Early Diagnosis of Alzheimer's Disease on MRI Images," *Neuroscience*, vol. 460, pp. 43–52, Apr. 2021, doi: 10.1016/j.neuroscience.2021.01.002.

[15] D. Agarwal, G. Marques, I. De La Torre-Díez, M. A. Franco Martin, B. García Zapiraín, and F. Martín Rodríguez, “Transfer Learning for Alzheimer’s Disease through Neuroimaging Biomarkers: A Systematic Review,” *Sensors*, vol. 21, no. 21, p. 7259, Oct. 2021, doi: 10.3390/s21217259.

[16] M. A. DeTure and D. W. Dickson, “The neuropathological diagnosis of Alzheimer’s disease,” *Mol. Neurodegener.*, vol. 14, no. 1, p. 32, Dec. 2019, doi: 10.1186/s13024-019-0333-5.

[17] R. R. Janghel and Y. K. Rathore, “Deep Convolution Neural Network Based System for Early Diagnosis of Alzheimer’s Disease,” *IRBM*, vol. 42, no. 4, pp. 258–267, Aug. 2021, doi: 10.1016/j.irbm.2020.06.006.

[18] X. Zhao, C. K. E. Ang, U. R. Acharya, and K. H. Cheong, “Application of Artificial Intelligence techniques for the detection of Alzheimer’s disease using structural MRI images,” *Biocybern. Biomed. Eng.*, vol. 41, no. 2, pp. 456–473, Apr. 2021, doi: 10.1016/j.bbe.2021.02.006.

[19] Y. Zhao *et al.*, “Application of Deep Learning for Prediction of Alzheimer’s Disease in PET/MR Imaging,” 2023.

[20] B. Kieffer, M. Babaie, S. Kalra, and H. R. Tizhoosh, “Convolutional neural networks for histopathology image classification: Training vs. Using pre-trained networks,” in *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*, Nov. 2017, pp. 1–6. doi: 10.1109/IPTA.2017.8310149.

[21] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, “Deep learning for visual understanding: A review,” *Neurocomputing*, vol. 187, pp. 27–48, Apr. 2016, doi: 10.1016/j.neucom.2015.09.116.

[22] P. P. Shinde and S. Shah, “A Review of Machine Learning and Deep Learning Applications,” in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, Pune, India: IEEE, Aug. 2018, pp. 1–6. doi: 10.1109/ICCUBEA.2018.8697857.

[23] “What Is Deep Learning? | How It Works, Techniques & Applications.” Accessed: Sep. 10, 2023. [Online]. Available: https://www.mathworks.com/discovery/deep-learning.html

[24] D. Pruthviraja *et al.*, “Detection of Alzheimer’s Disease Based on Cloud-Based Deep Learning Paradigm,” *Diagnostics*, vol. 13, no. 16, p. 2687, Aug. 2023, doi: 10.3390/diagnostics13162687.

[25] P. Grover *et al.*, “Ensemble Transfer Learning for Distinguishing Cognitively Normal and Mild Cognitive Impairment Patients Using MRI,” *Algorithms*, vol. 16, no. 8, p. 377, Aug. 2023, doi: 10.3390/a16080377.

[26] S. Naz, A. Ashraf, and A. Zaib, “Transfer learning using freeze features for Alzheimer neurological disorder detection using ADNI dataset,” *Multimed. Syst.*, vol. 28, no. 1, pp. 85–94, Feb. 2022, doi: 10.1007/s00530-021-00797-3.

[27] K. K. Parhi and N. K. Unnikrishnan, “Brain-Inspired Computing: Models and Architectures,” *IEEE Open J. Circuits Syst.*, vol. 1, pp. 185–204, 2020, doi: 10.1109/OJCAS.2020.3032092.

[28] N. M. Khan, N. Abraham, and M. Hon, “Transfer Learning With Intelligent Training Data Selection for Prediction of Alzheimer’s Disease,” *IEEE Access*, vol. 7, pp. 72726–72735, 2019, doi: 10.1109/ACCESS.2019.2920448.
[29] A. Loddo, S. Buttau, and C. Di Ruberto, “Deep learning based pipelines for Alzheimer’s disease diagnosis: A comparative study and a novel deep-ensemble method,” *Comput. Biol. Med.*, vol. 141, p. 105032, Feb. 2022, doi: 10.1016/j.compbiomed.2021.105032.
[30] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. in Adaptive computation and machine learning. Cambridge, Mass: The MIT press, 2016.
[31] Y. Xiao, S. Ma, J. Zhai, C. Fang, J. Jia, and Z. Chen, “Towards General Robustness Verification of MaxPool-Based Convolutional Neural Networks via Tightening Linear Approximation,” in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA: IEEE, Jun. 2024, pp. 24766–24775. doi: 10.1109/CVPR52733.2024.02339.
[32] S. R. Dubey, S. K. Singh, and B. B. Chaudhuri, “Activation functions in deep learning: A comprehensive survey and benchmark,” *Neurocomputing*, vol. 503, pp. 92–108, Sep. 2022, doi: 10.1016/j.neucom.2022.06.111.
[33] A. F. Agarap, “Deep Learning using Rectified Linear Units (ReLU),” Feb. 07, 2019, *arXiv*: arXiv:1803.08375. doi: 10.48550/arXiv.1803.08375.
[34] Z. Yi and Y. Wang, “Transfer Learning on Interstitial Lung Disease Classification,” in *2021 International Conference on Signal Processing and Machine Learning (CONF-SPML)*, Nov. 2021, pp. 199–205. doi: 10.1109/CONF-SPML54095.2021.00046.
[35] P. Ramachandran, B. Zoph, and Q. V. Le, “Searching for Activation Functions,” 2017, *arXiv*. doi: 10.48550/ARXIV.1710.05941.
[36] D. Hendrycks and K. Gimpel, “Gaussian Error Linear Units (GELUs),” 2016, *arXiv*. doi: 10.48550/ARXIV.1606.08415.
[37] X. Zhang, X. Zhou, M. Lin, and J. Sun, “ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT: IEEE, Jun. 2018, pp. 6848–6856. doi: 10.1109/CVPR.2018.00716.
[38] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” in *Proceedings of the 32nd International Conference on Machine Learning*, PMLR, Jun. 2015, pp. 448–456. Accessed: Feb. 01, 2025. [Online]. Available: https://proceedings.mlr.press/v37/ioffe15.html
[39] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” *J. Mach. Learn. Res.*, vol. 15, no. 56, pp. 1929–1958, 2014.
[40] J. Frankle and M. Carbin, “The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks,” 2018, doi: 10.48550/ARXIV.1803.03635.
[41] M. Lin, Q. Chen, and S. Yan, “Network In Network,” 2013, *arXiv*. doi: 10.48550/ARXIV.1312.4400.
[42] A. Vaswani *et al.*, “Attention Is All You Need,” 2017, *arXiv*. doi: 10.48550/ARXIV.1706.03762.

[43] T. N. Kipf and M. Welling, “Semi-Supervised Classification with Graph Convolutional Networks,” 2016, *arXiv*. doi: 10.48550/ARXIV.1609.02907.
[44] A. Howard *et al.*, “Searching for MobileNetV3,” in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South): IEEE, Oct. 2019, pp. 1314–1324. doi: 10.1109/ICCV.2019.00140.
[45] Z. Liu *et al.*, “Swin Transformer: Hierarchical Vision Transformer using Shifted Windows,” in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada: IEEE, Oct. 2021, pp. 9992–10002. doi: 10.1109/ICCV48922.2021.00986.
[46] G. Ayana, K. Dese, and S. Choe, “Transfer Learning in Breast Cancer Diagnoses via Ultrasound Imaging,” *Cancers*, vol. 13, no. 4, p. 738, Feb. 2021, doi: 10.3390/cancers13040738.
[47] S. Kornblith, J. Shlens, and Q. V. Le, “Do Better ImageNet Models Transfer Better?,” Jun. 17, 2019, *arXiv*: arXiv:1805.08974. doi: 10.48550/arXiv.1805.08974.
[48] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, “How transferable are features in deep neural networks?,” in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2014. Accessed: Feb. 01, 2025. [Online]. Available: https://proceedings.neurips.cc/paper/2014/hash/375c71349b295fbe2dcdca9206f20a06-Abstract.html
[49] S. Agrawal and P. Agrawal, “Detection of Pneumonia Cases from X-ray Chest Images using Deep Learning Based on Transfer Learning CNN and Hyperparameter Optimization,” in *2023 Third International Conference on Secure Cyber Computing and Communication (ICSCCC)*, Jalandhar, India: IEEE, May 2023, pp. 37–42. doi: 10.1109/ICSCCC58608.2023.10176853.
[50] K. He, R. Girshick, and P. Dollar, “Rethinking ImageNet Pre-Training,” in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2019, pp. 4917–4926. doi: 10.1109/ICCV.2019.00502.
[51] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Niessner, “FaceForensics++: Learning to Detect Manipulated Facial Images,” in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2019, pp. 1–11. doi: 10.1109/ICCV.2019.00009.
[52] S. Falkner, A. Klein, and F. Hutter, “BOHB: Robust and Efficient Hyperparameter Optimization at Scale,” in *Proceedings of the 35th International Conference on Machine Learning*, PMLR, Jul. 2018, pp. 1437–1446. Accessed: Feb. 02, 2025. [Online]. Available: https://proceedings.mlr.press/v80/falkner18a.html
[53] I. Loshchilov and F. Hutter, “Decoupled Weight Decay Regularization,” Jan. 04, 2019, *arXiv*: arXiv:1711.05101. doi: 10.48550/arXiv.1711.05101.
[54] F. Zhuang *et al.*, “A Comprehensive Survey on Transfer Learning,” *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021, doi: 10.1109/JPROC.2020.3004555.
[55] S. J. Pan and Q. Yang, “A Survey on Transfer Learning,” *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010, doi: 10.1109/TKDE.2009.191.
[56] S. Santurkar, A. Ilyas, D. Tsipras, L. Engstrom, B. Tran, and A. Madry, “Image Synthesis with a Single (Robust) Classifier,” in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2019. Accessed: Feb. 01, 2025. [Online]. Available:

https://proceedings.neurips.cc/paper/2019/hash/6f2268bd1d3d3ebaabb04d6b5d099425-Abstract.html
[57] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, “A Survey on Deep Transfer Learning,” in *Artificial Neural Networks and Machine Learning – ICANN 2018*, V. Kůrková, Y. Manolopoulos, B. Hammer, L. Iliadis, and I. Maglogiannis, Eds., Cham: Springer International Publishing, 2018, pp. 270–279. doi: 10.1007/978-3-030-01424-7_27.
[58] Y.-L. Li *et al.*, “Transferable Interactiveness Knowledge for Human-Object Interaction Detection,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2019, pp. 3580–3589. doi: 10.1109/CVPR.2019.00370.
[59] Y. Ganin *et al.*, “Domain-Adversarial Training of Neural Networks,” *J. Mach. Learn. Res.*, vol. 17, no. 59, pp. 1–35, 2016.
[60] B. Bischl *et al.*, “Hyperparameter Optimization: Foundations, Algorithms, Best Practices and Open Challenges,” Nov. 24, 2021, *arXiv*: arXiv:2107.05847. doi: 10.48550/arXiv.2107.05847.
[61] S. Kapoor and A. Narayanan, “Leakage and the reproducibility crisis in machine-learning-based science,” *Patterns*, vol. 4, no. 9, p. 100804, Sep. 2023, doi: 10.1016/j.patter.2023.100804.
[62] S. Kaufman, S. Rosset, and C. Perlich, “Leakage in data mining: formulation, detection, and avoidance,” in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, San Diego California USA: ACM, Aug. 2011, pp. 556–563. doi: 10.1145/2020408.2020496.
[63] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A Simple Framework for Contrastive Learning of Visual Representations,” Jul. 01, 2020, *arXiv*: arXiv:2002.05709. doi: 10.48550/arXiv.2002.05709.
[64] P. Refaeilzadeh, L. Tang, and H. Liu, “Cross-Validation,” in *Encyclopedia of Database Systems*, L. LIU and M. T. ÖZSU, Eds., Boston, MA: Springer US, 2009, pp. 532–538. doi: 10.1007/978-0-387-39940-9_565.
[65] T. Hidayat *et al.*, “Performance Prediction Using Cross Validation (GridSearchCV) for Stunting Prevalence,” in *2024 IEEE International Conference on Artificial Intelligence and Mechatronics Systems (AIMS)*, Feb. 2024, pp. 1–6. doi: 10.1109/AIMS61812.2024.10512657.
[66] M. Feurer, A. Klein, K. Eggensperger, J. T. Springenberg, M. Blum, and F. Hutter, “Efficient and robust automated machine learning,” in *Proceedings of the 29th International Conference on Neural Information Processing Systems - Volume 2*, in NIPS’15, vol. 2. Cambridge, MA, USA: MIT Press, Dec. 2015, pp. 2755–2763.
[67] S. Varma and R. Simon, “Bias in error estimation when using cross-validation for model selection,” *BMC Bioinformatics*, vol. 7, no. 1, p. 91, Feb. 2006, doi: 10.1186/1471-2105-7-91.
[68] “Machine Learning Testing: Survey, Landscapes and Horizons | IEEE Journals & Magazine | IEEE Xplore.” Accessed: Feb. 17, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/9000651

[69] D. Hendrycks *et al.*, “The Many Faces of Robustness: A Critical Analysis of Out-of-Distribution Generalization,” Jul. 24, 2021, *arXiv*: arXiv:2006.16241. doi: 10.48550/arXiv.2006.16241.

[70] A. E. Cinà *et al.*, “AttackBench: Evaluating Gradient-based Attacks for Adversarial Examples,” Oct. 03, 2024, *arXiv*: arXiv:2404.19460. doi: 10.48550/arXiv.2404.19460.

[71] R. Liaw, E. Liang, R. Nishihara, P. Moritz, J. E. Gonzalez, and I. Stoica, “Tune: A Research Platform for Distributed Model Selection and Training,” Jul. 13, 2018, *arXiv*: arXiv:1807.05118. doi: 10.48550/arXiv.1807.05118.

[72] H. Semmelrock *et al.*, “Reproducibility in Machine Learning-based Research: Overview, Barriers and Drivers,” Jul. 02, 2024, *arXiv*: arXiv:2406.14325. doi: 10.48550/arXiv.2406.14325.

[73] N. E. Khalifa, M. Loey, and S. Mirjalili, “A comprehensive survey of recent trends in deep learning for digital images augmentation,” *Artif. Intell. Rev.*, vol. 55, no. 3, pp. 2351–2377, Mar. 2022, doi: 10.1007/s10462-021-10066-4.

[74] M. Xu, S. Yoon, A. Fuentes, and D. S. Park, “A Comprehensive Survey of Image Augmentation Techniques for Deep Learning,” *Pattern Recognit.*, vol. 137, p. 109347, May 2023, doi: 10.1016/j.patcog.2023.109347.

[75] M. Xu, S. Yoon, A. Fuentes, and D. S. Park, “A Comprehensive Survey of Image Augmentation Techniques for Deep Learning,” *Pattern Recognit.*, vol. 137, p. 109347, May 2023, doi: 10.1016/j.patcog.2023.109347.

[76] M. A. Ali, F. Dornaika, A. Cherry, M. A. Ali, and A. Hage-Diab, “Advancing Medical Image Processing: A Comparative Analysis of ‘Naturalize’ and Traditional Augmentation Techniques for Enhanced Diagnostic Accuracy,” in *2024 International Conference on Smart Systems and Power Management (IC2SPM)*, Nov. 2024, pp. 138–143. doi: 10.1109/IC2SPM62723.2024.10841335.

[77] L. Nanni, M. Paci, S. Brahnam, and A. Lumini, “Comparison of Different Image Data Augmentation Approaches,” *J. Imaging*, vol. 7, no. 12, p. 254, Nov. 2021, doi: 10.3390/jimaging7120254.

[78] A. Mikolajczyk and M. Grochowski, “Data augmentation for improving deep learning in image classification problem,” in *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, Swinoujście: IEEE, May 2018, pp. 117–122. doi: 10.1109/IIPHDW.2018.8388338.

[79] R. Logan *et al.*, “Deep Convolutional Neural Networks With Ensemble Learning and Generative Adversarial Networks for Alzheimer’s Disease Image Data Classification,” *Front. Aging Neurosci.*, vol. 13, p. 720226, Aug. 2021, doi: 10.3389/fnagi.2021.720226.

[80] Y. Jing, Y. Yang, Z. Feng, J. Ye, Y. Yu, and M. Song, “Neural Style Transfer: A Review,” *IEEE Trans. Vis. Comput. Graph.*, vol. 26, no. 11, pp. 3365–3385, Nov. 2020, doi: 10.1109/TVCG.2019.2921336.

[81] J. Islam and Y. Zhang, “Brain MRI analysis for Alzheimer’s disease diagnosis using an ensemble system of deep convolutional neural networks,” *Brain Inform.*, vol. 5, no. 2, p. 2, Dec. 2018, doi: 10.1186/s40708-018-0080-3.

[82] M. R. Ahmed, Y. Zhang, Z. Feng, B. Lo, O. T. Inan, and H. Liao, “Neuroimaging and Machine Learning for Dementia Diagnosis: Recent Advancements and Future

Prospects," *IEEE Rev. Biomed. Eng.*, vol. 12, pp. 19–33, 2019, doi: 10.1109/RBME.2018.2886237.
[83] S. Ahmed *et al.*, "Ensembles of Patch-Based Classifiers for Diagnosis of Alzheimer Diseases," *IEEE Access*, vol. 7, pp. 73373–73383, 2019, doi: 10.1109/ACCESS.2019.2920011.
[84] S. Afzal *et al.*, "A Data Augmentation-Based Framework to Handle Class Imbalance Problem for Alzheimer's Stage Detection," *IEEE Access*, vol. 7, pp. 115528–115539, 2019, doi: 10.1109/ACCESS.2019.2932786.
[85] C. Jimenez-Mesa *et al.*, "Optimized One vs One Approach in Multiclass Classification for Early Alzheimer's Disease and Mild Cognitive Impairment Diagnosis," *IEEE Access*, vol. 8, pp. 96981–96993, 2020, doi: 10.1109/ACCESS.2020.2997736.
[86] M. Wang, C. Lian, D. Yao, D. Zhang, M. Liu, and D. Shen, "Spatial-Temporal Dependency Modeling and Network Hub Detection for Functional MRI Analysis via Convolutional-Recurrent Network," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 8, pp. 2241–2252, Aug. 2020, doi: 10.1109/TBME.2019.2957921.
[87] H. Guo and Y. Zhang, "Resting State fMRI and Improved Deep Learning Algorithm for Earlier Detection of Alzheimer's Disease," *IEEE Access*, vol. 8, pp. 115383–115392, 2020, doi: 10.1109/ACCESS.2020.3003424.
[88] D. Prakash, N. Madusanka, S. Bhattacharjee, C.-H. Kim, H.-G. Park, and H.-K. Choi, "Diagnosing Alzheimer's Disease Based on Multiclass MRI Scans usingTransfer Learning Techniques," *Curr. Med. Imaging Former. Curr. Med. Imaging Rev.*, vol. 17, no. 12, pp. 1460–1472, Dec. 2021, doi: 10.2174/1573405617666210127161812.
[89] T.-E. Kam, H. Zhang, Z. Jiao, and D. Shen, "Deep Learning of Static and Dynamic Brain Functional Networks for Early MCI Detection," *IEEE Trans. Med. Imaging*, vol. 39, no. 2, pp. 478–487, Feb. 2020, doi: 10.1109/TMI.2019.2928790.
[90] J. Zhang, B. Zheng, A. Gao, X. Feng, D. Liang, and X. Long, "A 3D densely connected convolution neural network with connection-wise attention mechanism for Alzheimer's disease classification," *Magn. Reson. Imaging*, vol. 78, pp. 119–126, May 2021, doi: 10.1016/j.mri.2021.02.001.
[91] N. J. Herzog and G. D. Magoulas, "Brain Asymmetry Detection and Machine Learning Classification for Diagnosis of Early Dementia," *Sensors*, vol. 21, no. 3, p. 778, Jan. 2021, doi: 10.3390/s21030778.
[92] K. Nemoto *et al.*, "Differentiating Dementia with Lewy Bodies and Alzheimer's Disease by Deep Learning to Structural MRI," *J. Neuroimaging*, vol. 31, no. 3, pp. 579–587, May 2021, doi: 10.1111/jon.12835.
[93] Q. Li and M. Q. Yang, "Comparison of machine learning approaches for enhancing Alzheimer's disease classification," *PeerJ*, vol. 9, p. e10549, Feb. 2021, doi: 10.7717/peerj.10549.
[94] M. Odusami, R. Maskeliūnas, R. Damaševičius, and T. Krilavičius, "Analysis of Features of Alzheimer's Disease: Detection of Early Stage from Functional Brain Changes in Magnetic Resonance Images Using a Finetuned ResNet18 Network," *Diagnostics*, vol. 11, no. 6, p. 1071, Jun. 2021, doi: 10.3390/diagnostics11061071.

[95] M. Odusami, R. Maskeliūnas, and R. Damaševičius, “An Intelligent System for Early Recognition of Alzheimer’s Disease Using Neuroimaging,” *Sensors*, vol. 22, no. 3, p. 740, Jan. 2022, doi: 10.3390/s22030740.
[96] Sachin Kumar and Sourabh Shastri, “Alzheimer MRI Preprocessed Dataset.” Kaggle. doi: 10.34740/KAGGLE/DSV/3364939.
[97] J. V. Shanmugam, B. Duraisamy, B. C. Simon, and P. Bhaskaran, “Alzheimer’s disease classification using pre-trained deep networks,” *Biomed. Signal Process. Control*, vol. 71, p. 103217, Jan. 2022, doi: 10.1016/j.bspc.2021.103217.
[98] T. Rahman *et al.*, “Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-ray,” *Appl. Sci.*, vol. 10, no. 9, p. 3233, May 2020, doi: 10.3390/app10093233.
[99] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.
[100] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.
[101] C. Szegedy *et al.*, “Going deeper with convolutions,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015, pp. 1–9. doi: 10.1109/CVPR.2015.7298594.
[102] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” Dec. 11, 2015, *arXiv*: arXiv:1512.00567. doi: 10.48550/arXiv.1512.00567.
[103] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, “SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size,” Nov. 2016, Accessed: Mar. 25, 2025. [Online]. Available: https://openreview.net/forum?id=S1xh5sYgx
[104] R. A. Alsemmeari, M. Yehia Dahab, B. Alturki, and A. A. Alsulami, “Priority Detector and Classifier Techniques Based on ML for the IoMT,” *Comput. Mater. Contin.*, vol. 76, no. 2, pp. 1853–1870, 2023, doi: 10.32604/cmc.2023.038589.
[105] A. Alqahtani, A. A. Alsulami, S. Alahmari, and M. Alrizq, “Accurate Phase Detection for ZigBee Using Artificial Neural Network,” *Intell. Autom. Soft Comput.*, vol. 36, no. 3, pp. 2505–2518, 2023, doi: 10.32604/iasc.2023.033243.
[106] P. Fergus and C. Chalmers, “Performance Evaluation Metrics,” in *Applied Deep Learning: Tools, Techniques, and Implementation*, P. Fergus and C. Chalmers, Eds., Cham: Springer International Publishing, 2022, pp. 115–138. doi: 10.1007/978-3-031-04420-5_5.

END